

## MEAN AND VARIANCE OF BALANCED PÓLYA URNS

SVANTE JANSON,\* *Uppsala University*

### Abstract

It is well-known that in a small Pólya urn, i.e., an urn where the second largest real part of an eigenvalue is at most half the largest eigenvalue, the distribution of the numbers of balls of different colours in the urn is asymptotically normal under weak additional conditions. We consider the balanced case, and then give asymptotics of the mean and the covariance matrix, showing that after appropriate normalization, the mean and covariance matrix converge to the mean and covariance matrix of the limiting normal distribution.

*Keywords:* Pólya urns; moment convergence

2010 Mathematics Subject Classification: Primary 60C05

Secondary 60F25

### 1. Introduction

A (generalized) Pólya urn contains balls of different colours. A ball is drawn at random from the urn, and is replaced by a set of balls that depends on the colour of the drawn ball. (Moreover, the replacement set may be random, with a distribution depending on the drawn colour). This is repeated an infinite number of times, and we are interested in the asymptotic composition of the urn. For details, and the assumptions used in the present paper, see Section 2; for the history of Pólya urns, see e.g. [24].

It is well-known, and proved under various conditions in a number of papers by a variety of authors (see e.g. [19, Theorems 3.22–3.24]), that the asymptotic behaviour depends on the eigenvalues of the *intensity matrix* of the urn, defined in (2.5) below, and in particular on the two largest (in real part) eigenvalues,  $\lambda_1$  and  $\lambda_2$ . If  $\operatorname{Re} \lambda_2 \leq \frac{1}{2}\lambda_1$  (a *small urn*), then, under some assumptions (including some version of irreducibility), the number of balls of a given colour is asymptotically normal, while if  $\operatorname{Re} \lambda_2 > \frac{1}{2}\lambda_1$  (a *large urn*), then this is not true: there are (again under some assumptions, and after suitable normalization) limits in distribution, but the limiting distributions have no simple description and are (typically, at least) not normal; furthermore, there may be oscillations so that suitable subsequences converge in distribution but the full sequence does not. Another difference is that for a small urn, the limit is independent of the initial state, and therefore independent of what happens in any fixed finite set of draws (i.e., the limit theorem is mixing; see [1, Proposition 2]), while for a large urn, on the contrary, there is an almost sure (a.s.) limit result and thus the limit is essentially determined by what happens early in the process.

For large urns, assuming that the urn is *balanced* (see Section 2), Pouyanne [29] proved a limit theorem which shows such an a.s. result and also shows convergence in  $L^p$  for any  $p$ , and thus convergence of all moments.

---

Received 31 March 2016; revision received 17 December 2019.

\* Postal address: Department of Mathematics, Uppsala University, PO Box 480, SE-751 06 Uppsala, Sweden. Email: [svante.janson@math.uu.se](mailto:svante.janson@math.uu.se)

For small urns, however, less has been known about moment convergence in general. Balanced deterministic urns with two colours were considered already by Bernstein [9, 10], who both showed the asymptotic normality in the small urn case and gave results on mean and variance; Savkevits [30] also considered this urn and studied the mean and variance and, moreover, the third and fourth moments. Bagchi and Pal [5] (independently, but 45 years later) gave another proof of asymptotic normality for balanced deterministic small urns with two colours, by using the method of moments and thus proving moment convergence as part of the proof. Bai and Hu [6, 7], who consider an arbitrary number of colours and allow random replacements (under somewhat different conditions than ours, allowing also time-dependent replacements), show asymptotic results for the mean and (co)variance as part of their proofs of asymptotic normality for small urns, using the same decomposition of the covariance matrix as in the present paper; however, these results are hidden inside the proof and are not stated explicitly. More recently, Janson and Pouyanne [21] proved explicit results on asymptotics of mean and variance, and also higher central moments of arbitrary order, for irreducible small urns; the method there combined the known result on asymptotic normality in this case, and moment estimates by the method in [29], leading to uniform integrability.

Note that asymptotics of mean and variance often are of interest in applications, complementing results on convergence in distribution. (In some applications, results on mean and variance have been proved separately by other methods.) Moreover, although not surprising, it is satisfying to know that in a case where a central limit theorem holds, the mean and variance also converge as suggested by this central limit theorem. In particular, loosely speaking, the variance of the number of balls of a given colour is asymptotic to the asymptotic variance. Furthermore, for this random number, the standard normalization by mean and standard deviation yields convergence to a standard normal distribution.

The main purpose of the present paper is to give explicit asymptotics for the first and second moments for a balanced small urn by an elementary direct method. (Our assumptions are somewhat weaker than in [21]; we do not assume that a central limit theorem holds, and also some non-normal cases are included; see Remark 3.2.) We also include a simple result on non-degeneracy of the limit (Theorem 3.5). Precise statements are given in Section 3. Some results (e.g. Theorem 3.1 and the lemmas in Sections 5 and 6) apply also to large urns.

Our method is closely related to the one used by e.g. [6, 7]; it is also related to the method of [29] and [21], but substantially simpler. The main idea (which has been used in various forms earlier, in particular by [6, 7] is that the drawing of a ball and the subsequent addition of a set of balls, at time  $k$ , say, influences the composition of the urn at a later time  $n$  not only directly by the added balls, but also indirectly since the added balls change the probabilities for later draws. By including the expectation of these later indirect effects, we find the real effect at time  $n$  of the draw at time  $k$ , and we may write the composition at time  $n$  as the sum, for  $k \leq n$ , of these contributions; see (4.11). The contributions for different  $k$  are orthogonal, and thus the variance can be found by summing the variances of these contributions.

Another purpose of the present paper is to demonstrate in detail this elementary method and how it can be used to obtain important results. We believe that although the method is closely related to earlier proofs in e.g. [6; 7; 17], making the decomposition (4.11) explicit illuminates both the proofs and the general behaviour of Pólya urns, in particular the difference between large and small urns. See the comments in Section 8.

Section 2 gives definitions and introduces the notation. Section 3 contains the statements of the main results, which are proved in Sections 4–6. Section 7 presents some applications, and Section 8 contains some further comments on where the variance comes from, i.e., which draws are most important, and the difference between small and large urns. The appendices give some further, more technical, results.

**Remark 1.1.** We consider in the present paper only the mean and (co)variance. As stated above, similar results on convergence of higher moments for balanced small urns are given in [21] (under somewhat more restrictive assumptions than in the present paper), by another method, based on showing uniform integrability. (An anonymous referee of the first version of the present paper has suggested another, simpler, way to show uniform integrability; this can be used to simplify the proofs in [21].)

It is possible that the method in the present paper can be extended to handle higher moments too, but we do not see any immediate extension. On the other hand, for the first and second moments, the present method seems simpler, and perhaps also more informative, than the one in [21].

**Problem 1.1.** In the present paper, we consider only balanced urns. We leave it as a challenging open problem to prove (or disprove?) similar results for non-balanced urns.

## 2. Pólya urns

### 2.1. Definition and assumptions

A (generalized) Pólya urn process is defined as follows. (See e.g. Mahmoud [24], Johnson and Kotz [23], Janson [19], Flajolet, Gabarró and Pekari [14] and Pouyanne [29] for the history and further references, as well as some different methods used to study such urns.) There are balls of  $q$  colours (types)  $1, \dots, q$ , where  $2 \leq q < \infty$ . The composition of the urn at time  $n$  is given by the vector  $X_n = (X_{n1}, \dots, X_{nq}) \in [0, \infty)^q$ , where  $X_{ni}$  is the number of balls of colour  $i$ . The urn starts with a given vector  $X_0$ , and evolves according to a discrete-time Markov process. Each colour  $i$  has an *activity* (or weight)  $a_i \geq 0$ , and a (generally random) *replacement vector*  $\xi_i = (\xi_{i1}, \dots, \xi_{iq})$ . At each time  $n + 1 \geq 1$ , the urn is updated by drawing one ball at random from the urn, with the probability of any ball proportional to its activity. Thus, the drawn ball has colour  $i$  with probability

$$\frac{a_i X_{ni}}{\sum_j a_j X_{nj}}. \tag{2.1}$$

If the drawn ball has type  $i$ , it is replaced together with  $\Delta X_{nj}$  balls of type  $j, j = 1, \dots, n$ , where the random vector  $\Delta X_n = (\Delta X_{n1}, \dots, \Delta X_{nq})$  has the same distribution as  $\xi_i$  and is independent of everything else that has happened so far. Thus, the urn is updated to  $X_{n+1} = X_n + \Delta X_n$ .

In many applications, the numbers  $X_{ni}$  and  $\xi_{ij}$  are integers, but that is not necessary; it has been noted several times that the Pólya urn process is well-defined also for *real*  $X_{ni}$  and  $\xi_{ij}$ , with probabilities for the different replacements still given by (2.1); see e.g. [6], [8, p. 126], [19, Remark 4.2], [20, Remark 1.11], and [29], and the earlier [22] for the related case of branching processes; the ‘number of balls’  $X_{ni}$  may thus be any non-negative real number. (This can be interpreted as the amount (mass) of colour  $i$  in the urn, rather than the number of discrete balls.) The replacements  $\xi_{ij}$  are thus random real numbers. We allow them to be negative, meaning that balls may be subtracted from the urn. However, we always assume that  $X_0$  and the random vectors  $\xi_i$  are such that, for every  $n \geq 0$ , a.s.

$$\text{each } X_{ni} \geq 0 \quad \text{and} \quad \sum_i a_i X_{ni} > 0, \tag{2.2}$$

so that (2.1) really gives meaningful probabilities (and the process does not stop due to lack of balls to be removed). An urn with such initial conditions and replacement rules is called *tenable*.

**Remark 2.1.** A sufficient condition for tenability, which often is assumed in other papers (sometimes with simple modifications), is that all  $\xi_{ij}$  and  $X_{0i}$  are integers with  $\xi_{ij} \geq 0$  for  $j \neq i$  and  $\xi_{ii} \geq -1$  (this means that we may remove the drawn ball but no other ball), and furthermore, for example,  $\sum_j a_j \xi_{ij} \geq 0$  a.s. (meaning that the total activity never decreases); then the urn is tenable for any  $X_0$  with non-zero activity. This is satisfied in most applications we know of, but not all; see Remark 7.1 for a different example. We shall *not* assume this condition in the present paper, unless explicitly stated.

**Remark 2.2.** In all applications that we know of, each  $\xi_i$  is a discrete random vector, i.e. it takes only a countable (usually a finite) number of different values. This is not necessary, however; the results below hold also if, e.g., some  $\xi_{ij}$  is continuous.

We assume, for simplicity, that the initial composition  $X_0$  is deterministic.

**Remark 2.3.** The results are easily extended to the case of random  $X_0$  by conditioning on  $X_0$ , but that may require some extra conditions or minor modifications in some of the statements, which we leave to the reader.

The Pólya urn is *balanced* if

$$\sum_j a_j \xi_{ij} = b > 0 \tag{2.3}$$

(a.s.) for some constant  $b$  and every  $i$ . In other words, the added activity after each draw is fixed (non-random and not depending on the colour of the drawn ball). This implies that the denominator in (2.1) (which is the total activity in the urn) is deterministic for each  $n$ ; see (4.9). This is a significant simplification, and is assumed in many papers on Pólya urns. (One exception is [19], which is based on embedding in a continuous-time branching process and stopping at a suitable stopping time, following [3]; this method does not seem to easily give information on moments and is not used in the present paper.)

**Remark 2.4.** We exclude the case  $b = 0$ , which is quite different; a typical example is a Markov chain, regarded as an urn always containing a single ball.

We shall assume that the urn is tenable and balanced; this is sometimes repeated for emphasis.

We also assume (2.9) below; as discussed in Remark 2.6 and Appendix A, this is a very weak assumption needed to exclude some trivial cases allowed by our definition of tenable; by Lemma A.1 it is sufficient to assume that every colour in the specification actually may occur in the urn, which always can be achieved by eliminating any redundant colours.

Finally, in order to obtain moment results, we assume that the replacements have second moments:

$$\mathbb{E} \xi_{ij}^2 < \infty, \quad i, j = 1, \dots, q. \tag{2.4}$$

It follows that every  $X_n$  has second moments, so the covariance matrix  $\text{Var}(X_n)$  is finite for each  $n$ .

**Remark 2.5.** In the tenable and balanced case, the assumption (2.4) is almost redundant. First, although there might be negative values of  $\xi_{ij}$ , we assume that the urn is tenable. Hence, given any instance  $(x_1, \dots, x_q)$  of the urn that may occur with positive probability as some  $X_n$ , we have  $\xi_{ij} \geq -x_j$  a.s. for every  $i$  and  $j$  such that  $a_i x_i > 0$ . In particular, if every  $a_i > 0$ , and every colour may appear in the urn, then each  $\xi_{ij}$  is bounded below. Furthermore, still assuming  $a_i > 0$  for each  $i$ , this and (2.3) implies that each  $\xi_{ij}$  also is bounded above; hence  $\xi_{ij}$  is bounded and has moments of any order.

**2.2. Notation**

We regard all vectors as column vectors. We use standard notation for (real or complex) vectors and matrices (of sizes  $q$  and  $q \times q$ , respectively), in particular  $'$  for transpose,  $*$  for Hermitian conjugate, and  $\cdot$  for the standard scalar product; thus  $u \cdot v = u'v$  for any vectors  $u, v \in \mathbb{R}^q$ . We let  $\| \cdot \|$  denote the standard Euclidean norm for vectors, and the operator norm (or any other convenient norm) for matrices.

Let  $a := (a_1, \dots, a_q)'$  be the vector of activities. Thus, the balance condition (2.3) can be written  $a \cdot \xi_i = b$ .

The *intensity matrix* of the Pólya urn is the  $q \times q$  matrix

$$A := (a_j \mathbb{E} \xi_{ji})_{i,j=1}^q. \tag{2.5}$$

(Note that, for convenience and following [19], we have defined  $A$  so that the element  $(A)_{ij}$  is a measure of the intensity of adding balls of colour  $i$  coming from drawn balls of colour  $j$ ; the transpose matrix  $A'$  is often used in other papers.) The intensity matrix  $A$  with its eigenvalues and eigenvectors has a central role for asymptotical results.

Let  $\sigma(A)$  (the *spectrum* of  $A$ ) be the set of eigenvalues of  $A$ .

We shall use the Jordan decomposition of the matrix  $A$  in the following form. There exists a decomposition of the complex space  $\mathbb{C}^q$  as a direct sum  $\bigoplus_{\lambda} E_{\lambda}$  of generalized eigenspaces  $E_{\lambda}$ , such that  $A - \lambda I$  is a nilpotent operator on  $E_{\lambda}$ ; here  $\lambda$  ranges over the set  $\sigma(A)$  of eigenvalues of  $A$ . ( $I$  is the identity matrix of appropriate size.) In other words, there exist projections  $P_{\lambda}$ ,  $\lambda \in \sigma(A)$ , that commute with  $A$  and satisfy

$$\sum_{\lambda \in \sigma(A)} P_{\lambda} = I, \tag{2.6}$$

$$AP_{\lambda} = P_{\lambda}A = \lambda P_{\lambda} + N_{\lambda}, \tag{2.7}$$

where  $N_{\lambda} = P_{\lambda}N_{\lambda} = N_{\lambda}P_{\lambda}$  is nilpotent. Moreover,  $P_{\lambda}P_{\mu} = 0$  when  $\lambda \neq \mu$ . We let  $v_{\lambda} \geq 0$  be the integer such that  $N_{\lambda}^{v_{\lambda}} \neq 0$  but  $N_{\lambda}^{v_{\lambda}+1} = 0$ . (Equivalently, in the Jordan normal form of  $A$ , the largest Jordan block with  $\lambda$  on the diagonal has size  $v_{\lambda} + 1$ .) Hence  $v_{\lambda} = 0$  if and only if  $N_{\lambda} = 0$ , and this happens for all  $\lambda$  if and only if  $A$  is diagonalizable, i.e. if and only if  $A$  has a complete set of  $q$  linearly independent eigenvectors. (In the sequel,  $\lambda$  will always denote an eigenvalue. We may for completeness define  $P_{\lambda} = N_{\lambda} = 0$  for every  $\lambda \notin \sigma(A)$ .)

The eigenvalues of  $A$  are denoted  $\lambda_1, \dots, \lambda_q$  (repeated according to their algebraic multiplicities); we assume that they are ordered with decreasing real parts,  $\text{Re } \lambda_1 \geq \text{Re } \lambda_2 \geq \dots$ , and furthermore, when the real parts are equal, in order of decreasing  $v_j := v_{\lambda_j}$ . In particular, if  $\lambda_1 > \text{Re } \lambda_2$ , then  $v_j \leq v_2$  for every eigenvalue  $\lambda_j$  with  $\text{Re } \lambda_j = \text{Re } \lambda_2$ .

Recall that the urn is called *small* if  $\text{Re } \lambda_2 \leq \frac{1}{2}\lambda_1$  and *large* if  $\text{Re } \lambda_2 > \frac{1}{2}\lambda_1$ ; the urn is *strictly small* if  $\text{Re } \lambda_2 < \frac{1}{2}\lambda_1$ .

In the balanced case, by (2.5) and (2.3),

$$a'A = \left( \sum_{i=1}^q a_i(A)_{ij} \right)_j = \left( \sum_{i=1}^q a_i a_j \mathbb{E} \xi_{ji} \right)_j = (a_j \mathbb{E}(a \cdot \xi_j))_j = ba'; \tag{2.8}$$

i.e.,  $a'$  is a left eigenvector of  $A$  with eigenvalue  $b$ . Thus  $b \in \sigma(A)$ . We shall assume that, moreover,  $b$  is the largest eigenvalue, i.e.,

$$\lambda_1 = b. \tag{2.9}$$

**Remark 2.6.** In fact, (2.9) is a very weak assumption. For example, if each  $\xi_{ij} \geq 0$ , then  $A$  is a matrix with non-negative elements, and since the eigenvector  $a'$  is non-negative, (2.9) is a consequence of the Perron–Frobenius theorem. The same holds (by considering  $A + cI$  for a suitable  $c > 0$ ) under the assumption in Remark 2.1. Under our, more general, definition of tenability, there are counterexamples (see Example A.1), but we show in Appendix A that they are so only in a trivial way, and that we may assume (2.9) without real loss of generality. (Of course, the proof of Lemma A.1, which uses Lemma 5.4, does not use the assumption (2.9).)

We shall in our theorems furthermore assume that  $\text{Re } \lambda_2 < \lambda_1$  (and often more), and thus that  $\lambda_1 = b$  is a simple eigenvalue. There are thus corresponding left and right eigenvectors  $u'_1$  and  $v_1$  that are unique up to normalization. By (2.8), we may choose  $u_1 = a$ . Furthermore, we let  $v_1$  be normalized by

$$u_1 \cdot v_1 = a \cdot v_1 = 1. \tag{2.10}$$

Then the projection  $P_{\lambda_1}$  is given by

$$P_{\lambda_1} = v_1 u'_1. \tag{2.11}$$

Consequently, in the balanced case, for any vector  $v \in \mathbb{R}^q$ ,

$$P_{\lambda_1} v = v_1 u'_1 v = v_1 a' v = (a \cdot v) v_1. \tag{2.12}$$

**Remark 2.7.** The dominant eigenvalue  $\lambda_1$  is simple, and  $\text{Re } \lambda_2 < \lambda_1$  if, for example, the matrix  $A$  is irreducible, but not in general. A simple counterexample is the original Pólya urn (see Markov [26], Eggenberger and Pólya [13], and Pólya [28]), where each ball is replaced together with  $b$  balls of the same colour (and every  $a_i = 1$ ); then  $A = bI$  and  $\lambda_1 = \dots = \lambda_q = b$ . As is well-known, the asymptotic behaviour is quite different in this case; in particular,  $X_n/n$  converges in distribution to a non-degenerate distribution and not to a constant; see e.g. [28] and [23].

Define also

$$\widehat{P} := \sum_{\lambda \neq \lambda_1} P_\lambda = I - P_{\lambda_1}. \tag{2.13}$$

Furthermore, define the symmetric matrix

$$B := \sum_{i=1}^q a_i v_{1i} \mathbb{E}(\xi_i \xi'_i), \tag{2.14}$$

and, if the urn is *strictly small*, noting that  $\widehat{P}$  commutes with  $e^{sA} := \sum_{k=0}^\infty (sA)^k/k!$ , define

$$\Sigma_I := \int_0^\infty \widehat{P} e^{sA} B e^{sA'} \widehat{P}' e^{-\lambda_1 s} ds. \tag{2.15}$$

This integral converges absolutely when the urn is strictly small, as can be seen from the proof of Theorem 3.2, or directly because  $\|\widehat{P} e^{sA}\| = O(s^{\nu_2} e^{\text{Re } \lambda_2 s})$  for  $s \geq 1$ , as is easily seen from Lemma 5.1. (The integral is matrix-valued; the space of  $q \times q$  matrices is a finite-dimensional space and the integral can be interpreted componentwise.) See also Appendix B.

Unspecified limits are as  $n \rightarrow \infty$ . As usual,  $a_n = O(b_n)$  means that  $a_n/b_n$  is bounded; here  $a_n$  may be vectors or matrices and  $b_n$  may be complex numbers; we do not insist that  $b_n$  be positive.

$[x]$  is the smallest integer greater than or equal to  $x$ .

### 3. Main results

Our main results on asymptotics of mean and variance are the following. Proofs are given in Section 6. As mentioned in the introduction, results of this type exist, mainly implicitly, in earlier work. In particular, under similar (but not identical) assumptions, Theorem 3.1 is implicit in Bai and Hu [7, Theorem 3.2], and its proof, and explicit in Pouyanne [29, Proposition 7.1]; Theorems 3.2 and 3.3 are implicit in Bai and Hu [6, 7].

**Theorem 3.1.** *If the Pólya urn is tenable and balanced, and  $\operatorname{Re} \lambda_2 < \lambda_1$ , then for  $n \geq 2$ ,*

$$\begin{aligned} \mathbb{E}X_n &= (n\lambda_1 + a \cdot X_0)v_1 + O(n^{\operatorname{Re} \lambda_2/\lambda_1} \log^{v_2} n) \\ &= n\lambda_1 v_1 + O(n^{\operatorname{Re} \lambda_2/\lambda_1} \log^{v_2} n + 1) \\ &= n\lambda_1 v_1 + o(n). \end{aligned} \tag{3.1}$$

*In particular, if the urn is strictly small, i.e.  $\operatorname{Re} \lambda_2 < \frac{1}{2}\lambda_1$ , then*

$$\mathbb{E}X_n = n\lambda_1 v_1 + o(n^{1/2}). \tag{3.2}$$

**Theorem 3.2.** *If the Pólya urn is tenable, balanced, and strictly small, i.e.  $\operatorname{Re} \lambda_2 < \frac{1}{2}\lambda_1$ , then*

$$n^{-1} \operatorname{Var}(X_n) \rightarrow \Sigma := \lambda_1 \Sigma_I. \tag{3.3}$$

**Theorem 3.3.** *If the Pólya urn is tenable, balanced, and small but not strictly small, i.e.  $\operatorname{Re} \lambda_2 = \frac{1}{2}\lambda_1$ , then*

$$(n \log^{2v_2+1} n)^{-1} \operatorname{Var}(X_n) \rightarrow \frac{\lambda_1^{-2v_2}}{(2v_2 + 1)(v_2!)^2} \sum_{\operatorname{Re} \lambda = \frac{1}{2}\lambda_1} N_\lambda^{v_2} P_\lambda B P_\lambda^* (N_\lambda^*)^{v_2}.$$

**Remark 3.1.** Under some additional assumptions (irreducibility of  $A$ , at least if we ignore colours with activity 0, and, for example, the condition in Remark 2.1), [19, Theorems 3.22–3.23 and Lemma 5.4], show that if the urn is small, then  $X_n$  is asymptotically normal, with the asymptotic covariance matrix equal to the limit in Theorem 3.2 ( $\operatorname{Re} \lambda_2 < \frac{1}{2}\lambda_1$ ) or Theorem 3.3 ( $\operatorname{Re} \lambda_2 = \frac{1}{2}\lambda_1$ ). For example, in the strictly small case,  $n^{-1/2}(X_n - n\lambda_1 v_1) \xrightarrow{d} N(0, \Sigma)$ . Hence (under these hypotheses), Theorems 3.1–3.3 can be summarized by saying that the mean and (co)variances converge as expected in these central limit theorems.

We also obtain the following version of the law of large numbers for Pólya urns. Convergence a.s. has been shown before under various assumptions, including the unbalanced case as well as time-inhomogeneous generalizations (see [4, Section V.9.3], [2], [8], [18, Theorem 2.2], [19, Theorem 3.21], and [7, Theorem 2.2]), and is included here for completeness and because our conditions are somewhat more general. The  $L^2$  result is in [29, Remark 7.1(2)].

**Theorem 3.4.** *If the Pólya urn is tenable and balanced, and  $\operatorname{Re} \lambda_2 < \lambda_1$ , then as  $n \rightarrow \infty$ ,  $X_n/n \rightarrow \lambda_1 v_1$  a.s. and in  $L^2$ .*

The asymptotic covariance matrix  $\Sigma$  in (3.3) is always singular, since, by (2.15),  $\Sigma = \widehat{P}\Sigma\widehat{P}'$  and thus  $u'\Sigma u = u'\widehat{P}\Sigma\widehat{P}'u = 0$  when  $\widehat{P}'u = 0$ , which happens when  $P'_{\lambda_1} u = u$ , i.e., when  $u$  is



a multiple of the left eigenvector  $u_1 = a$ . In the balanced case, this is easy to see:  $a \cdot X_n$  is deterministic and thus  $\text{Var}(a \cdot X_n) = 0$ ; hence  $a' \Sigma a = 0$ , since for any vector  $u$ , by (3.3),

$$n^{-1} \text{Var}(u \cdot X_n) = n^{-1} u' \text{Var}(X_n) u \rightarrow u' \Sigma u. \tag{3.4}$$

With an extra assumption, this is the only case when the asymptotic variance  $u' \Sigma u$  vanishes (cf. [19, Remark 3.19]). Let  $\tilde{A}$  be the submatrix of  $A$  obtained by deleting all rows and columns corresponding to colours with activity  $a_i = 0$ .

**Theorem 3.5.** *Suppose that the Pólya urn is tenable, balanced, and strictly small, i.e.  $\text{Re } \lambda_2 < \frac{1}{2} \lambda_1$ , and, furthermore, that  $\tilde{A}$  is irreducible. If  $u \in \mathbb{R}^q$ , then  $u' \Sigma u = 0$  if and only if for every  $n \geq 0$ ,  $\text{Var}(u \cdot X_n) = 0$ , i.e.,  $u \cdot X_n$  is deterministic.*

**Remark 3.2.** If  $\tilde{A}$  is reducible, then, on the contrary,  $\Sigma$  is typically more singular. As an extreme example, consider a ‘triangular’ urn with two colours, activities  $a_i = 1$ , and deterministic replacements  $\xi_1 = (1, 0)$ ,  $\xi_2 = (1 - \lambda, \lambda)$  for a real  $\lambda \in (0, 1)$  (starting with one ball of each colour, say). Then  $A = \begin{pmatrix} 1 & 1-\lambda \\ 0 & \lambda \end{pmatrix}$ . The eigenvalues are 1 and  $\lambda$ , so the urn is strictly small if  $\lambda < \frac{1}{2}$ . However,  $v_1 = (1, 0)$ , and thus (2.14) yields  $B = \xi_1 \xi_1' = v_1 v_1'$ , and thus by (B.4) (or a direct calculation)  $\widehat{P}B = 0$ , and thus  $\Sigma = \Sigma_I = 0$ . Theorems 3.2 and 3.3 are still valid, but say only that the limit is 0. In fact, in this example, the proper normalization is  $n^\lambda$ : it follows from [20, Theorem 1.3(v)] that  $n^{-\lambda} X_{n2} = n^{-\lambda} (n + 2 - X_{n1}) \xrightarrow{d} W$  for some non-degenerate (and non-normal) random variable  $W$ . Moreover, calculations similar to those in Section 6 show that  $\mathbb{E} X_{n2} \sim c_1 n^\lambda$  and  $\text{Var} X_{n2} \sim c_2 n^{2\lambda}$  for some  $c_1, c_2 > 0$ , as shown earlier in [29, Example 7.2(2)].

**Remark 3.3.** It is easily seen that  $\tilde{A}$  is irreducible if and only if  $v_{1i} > 0$  for every  $i$  with  $a_i > 0$ .

### 4. Proofs, first steps

Let  $I_n$  be the colour of the  $n$ th drawn ball, let

$$\Delta X_n := X_{n+1} - X_n, \tag{4.1}$$

and let

$$w_n := a \cdot X_n, \tag{4.2}$$

the total weight (activity) of the urn. Furthermore, let  $\mathcal{F}_n$  be the  $\sigma$ -field generated by  $X_1, \dots, X_n$ . Then, by the definition of the urn,

$$\mathbb{P}(I_{n+1} = j \mid \mathcal{F}_n) = \frac{a_j X_{nj}}{w_n}, \tag{4.3}$$

and, consequently, recalling (2.5),

$$\begin{aligned} \mathbb{E}(\Delta X_n \mid \mathcal{F}_n) &= \sum_{j=1}^q \mathbb{P}(I_{n+1} = j \mid \mathcal{F}_n) \mathbb{E} \xi_j = \frac{1}{w_n} \sum_{j=1}^q a_j X_{nj} \mathbb{E} \xi_j \\ &= \frac{1}{w_n} \left( \sum_{j=1}^q (A)_{ij} X_{nj} \right)_i = \frac{1}{w_n} A X_n. \end{aligned} \tag{4.4}$$

Define

$$Y_n := \Delta X_{n-1} - \mathbb{E}(\Delta X_{n-1} \mid \mathcal{F}_{n-1}). \tag{4.5}$$



Then  $Y_n$  is  $\mathcal{F}_n$ -measurable, and obviously,

$$\mathbb{E}(Y_n | \mathcal{F}_{n-1}) = 0 \tag{4.6}$$

and, by (4.1), (4.5) and (4.4),

$$X_{n+1} = X_n + Y_{n+1} + w_n^{-1}AX_n = (I + w_n^{-1}A)X_n + Y_{n+1}. \tag{4.7}$$

Consequently, by induction, for any  $n \geq 0$ ,

$$X_n = \prod_{k=0}^{n-1} (I + w_k^{-1}A)X_0 + \sum_{\ell=1}^n \prod_{k=\ell}^{n-1} (I + w_k^{-1}A)Y_\ell, \tag{4.8}$$

where (as below) an empty matrix product is interpreted as  $I$ .

We now use the assumption that the urn is balanced, so  $a \cdot \Delta X_n = b$ , and thus by (4.1)–(4.2),  $w_n$  is deterministic with

$$w_n = w_0 + nb, \tag{4.9}$$

where the initial weight  $w_0 = a \cdot X_0$ . We define the matrix products

$$F_{i,j} := \prod_{i \leq k < j} (I + w_k^{-1}A), \quad 0 \leq i \leq j, \tag{4.10}$$

and write (4.8) as

$$X_n = F_{0,n}X_0 + \sum_{\ell=1}^n F_{\ell,n}Y_\ell. \tag{4.11}$$

As said in the introduction, we can regard the term  $F_{\ell,n}Y_\ell$  as the real effect on  $X_n$  of the  $\ell$ th draw, including the expected later indirect effects.

Taking the expectation, since  $\mathbb{E}Y_\ell = 0$  by (4.6) and the  $F_{i,j}$  and  $X_0$  are non-random, we find

$$\mathbb{E}X_n = F_{0,n}X_0. \tag{4.12}$$

Hence, (4.11) can also be written

$$X_n - \mathbb{E}X_n = \sum_{\ell=1}^n F_{\ell,n}Y_\ell. \tag{4.13}$$

Consequently, the covariance matrix can be computed as

$$\begin{aligned} \text{Var}(X_n) &:= \mathbb{E}((X_n - \mathbb{E}X_n)(X_n - \mathbb{E}X_n)') \\ &= \mathbb{E} \sum_{i=1}^n \sum_{j=1}^n (F_{i,n}Y_i)(F_{j,n}Y_j)' \\ &= \sum_{i=1}^n \sum_{j=1}^n F_{i,n} \mathbb{E}(Y_i Y_j') F_{j,n}'. \end{aligned} \tag{4.14}$$

However, if  $i > j$ , then  $\mathbb{E}(Y_i | \mathcal{F}_j) = 0$  by (4.6), and since  $Y_j$  is  $\mathcal{F}_j$ -measurable, we have

$$\mathbb{E}(Y_i Y_j') = \mathbb{E}(\mathbb{E}(Y_i | \mathcal{F}_j) Y_j') = 0. \tag{4.15}$$

Taking the transpose we see that  $\mathbb{E}(Y_i Y_j') = 0$  also when  $i < j$ . Hence, all non-diagonal terms vanish in (4.14), and we find

$$\text{Var}(X_n) = \sum_{i=1}^n F_{i,n} \mathbb{E}(Y_i Y_i') F_{i,n}' \tag{4.16}$$

The formulas (4.12) and (4.16) form the basis of our proofs, and it remains mainly to analyse the matrix products  $F_{i,j}$ .

**Remark 4.1.** The formula (4.8) holds for general Pólya urns, also when they are not balanced. However, in the general case, the total weights  $w_k$  are random, and they are dependent on each other and on the  $Y_\ell$ , and it seems difficult to draw any useful consequences from (4.8); certainly the arguments above fail because the  $F_{i,j}$  would be random.

**Remark 4.2.** As remarked by a referee, since  $P_{\lambda_1} Y_\ell = 0$  by Lemma 6.1, we may also write (4.13) as

$$X_n - \mathbb{E}X_n = \sum_{\ell=1}^n F_{\ell,n} \widehat{P} Y_\ell = \sum_{\ell=1}^n \widehat{F}_{\ell,n} Y_\ell, \tag{4.17}$$

where  $\widehat{F}_{i,j} := \widehat{P} F_{i,j} = F_{i,j} \widehat{P} = \prod_{i \leq k < j} (I + w_k^{-1} \widehat{A}) - P_{\lambda_1}$  with  $\widehat{A} := \widehat{P} A$ . This could be used instead of (4.13) to make another version of the proofs below; the two versions are very similar and essentially equivalent. (See [18] for a version essentially of this type.) The form (4.17) has the advantage that we have eliminated the (large) deterministic part corresponding to  $P_{\lambda_1}$ ; for example, assuming  $b = 1$ , for  $1 \leq \ell \leq n$  we obtain  $\|\widehat{F}_{\ell,n}\| = O((n/\ell)^{\text{Re } \lambda_2} (1 + \log(n/\ell))^{\nu_2})$ ; see Lemma 5.5. Nevertheless, we prefer to use  $F_{i,j}$  in the proofs below.

### 5. Estimates of matrix functions

In this section we derive some estimates of  $F_{i,n}$ ; these are used in the next section together with (4.16) to obtain the variance asymptotics. The estimates of  $F_{i,n}$  are obtained by standard matrix calculus, including a Jordan decomposition of  $A$ . Similar estimates have been used in several related papers, e.g. [18], [7], [31], and [19]. For completeness we nevertheless give detailed proofs.

For notational convenience, we make from now on the simplifying assumption  $b = 1$ . (For emphasis and clarity, we repeat this assumption in some statements; it will always be in force, whether stated or not.) This is no loss of generality: we can divide all activities by  $b$  and let the new activities be  $\hat{a} := a/b$ ; this defines the same random evolution of the urn, and we have  $\hat{a} \cdot \xi_i = b/b = 1$  for every  $i$ , so the modified urn is also balanced, with balance  $\hat{b} = 1$ . Furthermore, the intensity matrix  $A$  in (2.5) is divided by  $b$ , so all eigenvalues  $\lambda_i$  are divided by  $b$ , but their ratios remain the same; the projections  $P_\lambda$  remain the same while the nilpotent parts  $N_\lambda$  are divided by  $b$ , and in both cases the indices are shifted; also, with the normalization (2.10),  $u_1 = a$  is divided by  $b$  while  $v_1$  is multiplied by  $b$ . It is now easy to check that  $\lambda_1 v_1$ ,  $B$ , and  $\lambda_1 \Sigma_I$  are invariant, and thus the theorems all follow from the special case  $b = 1$ . By the assumption (2.9) (see Remark 2.6 and Appendix A), we thus have  $\lambda_1 = 1$ .

Note that (4.9) now becomes

$$w_n = n + w_0. \tag{5.1}$$

Note also that (4.10) can be written  $F_{i,j} = f_{i,j}(A)$ , where  $0 \leq i \leq j$  and  $f_{i,j}$  is the polynomial

$$\begin{aligned}
 f_{i,j}(z) &:= \prod_{i \leq k < j} (1 + w_k^{-1}z) = \prod_{i \leq k < j} \frac{w_k + z}{w_k} = \prod_{i \leq k < j} \frac{k + w_0 + z}{k + w_0} \\
 &= \frac{\Gamma(j + w_0 + z) / \Gamma(i + w_0 + z)}{\Gamma(j + w_0) / \Gamma(i + w_0)} = \frac{\Gamma(j + w_0 + z)}{\Gamma(j + w_0)} \cdot \frac{\Gamma(i + w_0)}{\Gamma(i + w_0 + z)}.
 \end{aligned}
 \tag{5.2}$$

Recall that by the functional calculus in spectral theory (see e.g. [12, Chapter VII.1–3]), we can define  $f(A)$  not only for polynomials  $f(z)$  but for any function  $f(z)$  that is analytic in a neighbourhood of the spectrum  $\sigma(A)$ . Furthermore, if  $K$  is a compact set that contains  $\sigma(A)$  in its interior (for example a sufficiently large disc), then there exists a constant  $C$  (depending on  $A$  and  $K$ ) such that for every  $f$  analytic in a neighbourhood of  $K$ ,

$$\|f(A)\| \leq C \sup_{z \in K} |f(z)|.
 \tag{5.3}$$

We shall use the functional calculus mainly for polynomials and the entire functions  $z \mapsto t^z = e^{(\log t)z}$  for fixed  $t > 0$ ; in these cases,  $f(A)$  can be defined by a Taylor series expansion as was done before (2.15). Note also that the general theory applies to operators in a Banach space; we only need the simpler finite-dimensional case discussed in [12, Chapter VII.1].

We shall use the following formula for  $f(A)$ , where  $f^{(m)}$  denotes the  $m$ th derivative of  $f$ . (The formula can be seen as a Taylor expansion; see the proof.)

**Lemma 5.1.** *For any entire function  $f(\lambda)$ , and any  $\lambda \in \sigma(A)$ ,*

$$f(A)P_\lambda = \sum_{m=0}^{\nu_\lambda} \frac{1}{m!} f^{(m)}(\lambda) N_\lambda^m P_\lambda.
 \tag{5.4}$$

*Proof.* This is a standard formula in the finite-dimensional case (see [12, Theorem VII.1.8]), but we give for completeness a simple (and perhaps informative) proof when  $f$  is a polynomial (which is the only case that we use, and which furthermore implies the general case by [12, Theorem VII.1.5(d)]). We then have the Taylor expansion  $f(\lambda + z) = \sum_{m=0}^\infty \frac{1}{m!} f^{(m)}(\lambda) z^m$ , which can be seen as an algebraic identity for polynomials in  $z$  (the sum is really finite since  $f^{(m)} = 0$  for large  $m$ ), and thus

$$f(A)P_\lambda = f(\lambda I + N_\lambda)P_\lambda = \sum_{m=0}^\infty \frac{1}{m!} f^{(m)}(\lambda) N_\lambda^m P_\lambda,
 \tag{5.5}$$

where  $N_\lambda^m = 0$  when  $m > \nu_\lambda$ . □

Our strategy is to first show estimates for the polynomials  $f_{i,j}(z)$  in (5.2) and then use these together with (5.3) and (5.4) to show the estimates for  $F_{i,j} = f_{i,j}(A)$  that we need.

**Lemma 5.2.**

(i) *For every fixed  $i$ , as  $j \rightarrow \infty$ ,*

$$f_{i,j}(z) = j^z \frac{\Gamma(i + w_0)}{\Gamma(i + w_0 + z)} (1 + o(1)),
 \tag{5.6}$$

*uniformly for  $z$  in any fixed compact set in the complex plane.*

(ii) As  $i, j \rightarrow \infty$  with  $i \leq j$ ,

$$f_{i,j}(z) = j^z i^{-z} (1 + o(1)), \tag{5.7}$$

uniformly for  $z$  in any fixed compact set in the complex plane.

*Proof.* Both parts follow from (5.2) and the fact that

$$\frac{\Gamma(x+z)}{\Gamma(x)} = x^z (1 + o(1)), \tag{5.8}$$

uniformly for  $z$  in a compact set, as  $x \rightarrow \infty$  (with  $x$  real, say), which is an easy and well-known consequence of Stirling’s formula; see [27, 5.11.12]. (Note that  $\Gamma(i + w_0)/\Gamma(i + w_0 + z)$  is an entire function for any  $i \geq 0$ , since  $w_0 > 0$ .  $\Gamma(j + w_0 + z)/\Gamma(j + w_0)$  has poles, but for  $z$  in a fixed compact set, this function is analytic when  $j$  is large enough.)  $\square$

For the derivatives  $f_{i,j}^{(m)}(z)$  there are corresponding estimates.

**Lemma 5.3.** *Let  $m \geq 0$ .*

(i) For every fixed  $i \geq 0$ , as  $j \rightarrow \infty$ ,

$$f_{i,j}^{(m)}(z) = j^z (\log j)^m \frac{\Gamma(i + w_0)}{\Gamma(i + w_0 + z)} + o(j^z \log^m j), \tag{5.9}$$

uniformly for  $z$  in any fixed compact set in the complex plane.

(ii) As  $i, j \rightarrow \infty$  with  $i \leq j$ ,

$$f_{i,j}^{(m)}(z) = \left(\frac{j}{i}\right)^z \left(\log \frac{j}{i}\right)^m + o\left(\left(\frac{j}{i}\right)^z \left(1 + \log \frac{j}{i}\right)^m\right), \tag{5.10}$$

uniformly for  $z$  in any fixed compact set in the complex plane.

*Proof.* (i): Let  $g_j(z) = j^{-z} f_{i,j}(z)$ . Then, by (5.6),

$$g_j(z) = \frac{\Gamma(i + w_0)}{\Gamma(i + w_0 + z)} (1 + o(1)) = O(1) \quad \text{as } j \rightarrow \infty, \tag{5.11}$$

uniformly in each compact set, and thus by Cauchy’s estimates, for any  $\ell \geq 1$ ,

$$g_j^{(\ell)}(z) = O(1) \quad \text{as } j \rightarrow \infty, \tag{5.12}$$

uniformly in each compact set. By Leibniz’s rule,

$$\begin{aligned} f_{i,j}^{(m)}(z) &= \frac{d^m}{dz^m} (j^z g_j(z)) = \sum_{\ell=0}^m \binom{m}{\ell} \frac{d^\ell}{dz^\ell} j^z \cdot g_j^{(m-\ell)}(z) \\ &= \sum_{\ell=0}^m \binom{m}{\ell} (\log j)^\ell j^z g_j^{(m-\ell)}(z), \end{aligned} \tag{5.13}$$

and (5.9) follows by (5.11)–(5.12).

(ii): Similarly, for  $1 \leq i \leq j$ , let  $h_{i,j}(z) = (i/j)^z f_{i,j}(z)$ . Then, by (5.7),

$$h_{i,j}(z) = 1 + o(1) \quad \text{as } i, j \rightarrow \infty, \tag{5.14}$$

uniformly in each compact set, and thus by Cauchy’s estimates, for any  $\ell \geq 1$ ,

$$h_{i,j}^{(\ell)}(z) = \frac{d^\ell}{dz^\ell}(h_{i,j}(z) - 1) = o(1) \quad \text{as } i, j \rightarrow \infty, \tag{5.15}$$

uniformly in each compact set. By Leibniz’s rule,

$$\begin{aligned} f_{i,j}^{(m)}(z) &= \frac{d^m}{dz^m}((j/i)^z h_{i,j}(z)) = \sum_{\ell=0}^m \binom{m}{\ell} \frac{d^\ell}{dz^\ell}(j/i)^z \cdot h_{i,j}^{(m-\ell)}(z) \\ &= \sum_{\ell=0}^m \binom{m}{\ell} (\log(j/i))^\ell (j/i)^z h_{i,j}^{(m-\ell)}(z), \end{aligned} \tag{5.16}$$

and (5.10) follows by (5.14)–(5.15). □

We now apply these estimates to  $F_{i,j}$ , noting that by Lemma 5.1,

$$F_{i,j}P_\lambda = f_{i,j}(A)P_\lambda = \sum_{m=0}^{v_\lambda} \frac{1}{m!} f_{i,j}^{(m)}(\lambda) N_\lambda^m P_\lambda. \tag{5.17}$$

**Lemma 5.4.** *If  $b = 1$ , then, for  $n \geq 2$  and  $\lambda \in \sigma(A)$ ,*

$$F_{0,n}P_\lambda = n^\lambda \log^{v_\lambda} n \frac{\Gamma(w_0)}{v_\lambda! \Gamma(w_0 + \lambda)} N_\lambda^{v_\lambda} P_\lambda + o(n^{\operatorname{Re} \lambda} \log^{v_\lambda} n). \tag{5.18}$$

*Proof.* By (5.17) and (5.9),

$$F_{0,n}P_\lambda = \sum_{m=0}^{v_\lambda} \frac{1}{m!} f_{0,n}^{(m)}(\lambda) N_\lambda^m P_\lambda = \frac{1}{v_\lambda!} f_{0,n}^{(v_\lambda)}(\lambda) N_\lambda^{v_\lambda} P_\lambda + \sum_{m=0}^{v_\lambda-1} O(n^\lambda \log^m n), \tag{5.19}$$

which yields (5.18) by another application of (5.9). □

**Lemma 5.5.** *If  $b = 1$ , then, for  $1 \leq i \leq j$  and  $\lambda \in \sigma(A)$ ,*

$$F_{i,j}P_\lambda = O((j/i)^{\operatorname{Re} \lambda} (1 + \log(j/i))^{v_\lambda}). \tag{5.20}$$

*More precisely, for any  $v \geq v_\lambda$ , as  $i, j \rightarrow \infty$  with  $i \leq j$ ,*

$$\begin{aligned} F_{i,j}P_\lambda &= \frac{1}{v!} \left(\frac{j}{i}\right)^\lambda \log^v \left(\frac{j}{i}\right) N_\lambda^v P_\lambda + o\left(\left(\frac{j}{i}\right)^{\operatorname{Re} \lambda} \log^v \left(\frac{j}{i}\right)\right) \\ &\quad + O\left(\left(\frac{j}{i}\right)^{\operatorname{Re} \lambda} \left(1 + \log^{v-1} \left(\frac{j}{i}\right)\right)\right). \end{aligned} \tag{5.21}$$

*Proof.* This is similar to the proof of Lemma 5.4. First, (5.20) follows directly from (5.17) and (5.10).

For (5.21), note that the summation in (5.17) may be extended to  $m \leq v$ , since  $N_\lambda^m = 0$  when  $m > v_\lambda$ . Then use (5.10) for each term  $m = v$ . □

**Lemma 5.6.** *If  $\operatorname{Re} \lambda_2 < \lambda_1 = b = 1$ , then for  $0 \leq i \leq j$ ,*

$$F_{i,j}P_{\lambda_1} = f_{i,j}(\lambda_1)P_{\lambda_1} = \frac{j + w_0}{i + w_0} P_{\lambda_1}. \tag{5.22}$$

*Proof.* Since  $\lambda_1$  thus is assumed to be a simple eigenvalue,  $v_{\lambda_1} = 0$ . (Alternatively, see Lemma A.1.) Hence, (5.17) yields  $F_{i,j}P_{\lambda_1} = f_{i,j}(\lambda_1)P_{\lambda_1}$ . Furthermore, (5.2) yields

$$f_{i,j}(\lambda_1) = f_{i,j}(1) = \frac{j + w_0}{i + w_0}, \tag{5.23}$$

and (5.22) follows. □

**Lemma 5.7.** *For any fixed  $x \in (0, 1]$ , as  $n \rightarrow \infty$ ,*

$$F_{\lceil xn \rceil, n} \rightarrow x^{-A}. \tag{5.24}$$

*Proof.* Let  $K$  be a compact set containing  $\sigma(A)$  in its interior. As  $n \rightarrow \infty$ , by (5.7),

$$f_{\lceil xn \rceil, n}(z) = \left( \frac{n}{\lceil xn \rceil} \right)^z (1 + o(1)) = x^{-z} (1 + o(1)) = x^{-z} + o(1), \tag{5.25}$$

uniformly for  $z \in K$ . Consequently,  $f_{\lceil xn \rceil, n}(z) - x^{-z} \rightarrow 0$  uniformly on  $K$ , and thus  $F_{\lceil xn \rceil, n} - x^{-A} \rightarrow 0$  by (5.3). □

**Lemma 5.8.** *There exist  $i_0$  and  $C$  such that if  $i_0 \leq i \leq j \leq 2i$ , then  $\|F_{i,j}\|^{-1} \leq C$ .*

*Proof.* Again let  $K$  be a compact set containing  $\sigma(A)$  in its interior. By (5.7), we may choose  $i_0$  such that if  $i_0 \leq i \leq j$ , then  $|f_{i,j}(z)| \geq \frac{1}{2} |(j/i)^z|$  on  $K$ . If furthermore  $i \leq j \leq 2i$ , this implies  $|f_{i,j}(z)| \geq c$  on  $K$ , for some  $c > 0$ , and thus  $|f_{i,j}^{-1}(z)| \leq c^{-1}$  on  $K$ . The result follows by (5.3). (The condition  $j \leq 2i$  is not needed when  $\sigma(A) \subset \{\text{Re}z > 0\}$ , so we may assume  $\text{Re}z \geq 0$  for  $z \in K$ .) □

### 6. Completions of the proofs

*Proof of Theorem 3.1.* By (4.12) and (2.6),

$$\mathbb{E}X_n = \sum_{\lambda \in \sigma(A)} F_{0,n}P_{\lambda}X_0. \tag{6.1}$$

For each eigenvalue  $\lambda \neq \lambda_1$ , Lemma 5.4 shows that

$$F_{0,n}P_{\lambda}X_0 = O(n^{\text{Re } \lambda} \log^{v_{\lambda}} n) = O(n^{\text{Re } \lambda_2} \log^{v_2} n). \tag{6.2}$$

Furthermore, by (2.12),

$$P_{\lambda_1}X_0 = (a \cdot X_0)v_1 = w_0v_1, \tag{6.3}$$

and it follows from (5.22) that

$$F_{0,n}P_{\lambda_1}X_0 = \frac{n + w_0}{w_0} P_{\lambda_1}X_0 = \frac{n + w_0}{w_0} w_0v_1 = (n + w_0)v_1. \tag{6.4}$$

The result (3.1) follows (when  $\lambda_1 = 1$ ) from (6.1), (6.2), and (6.4). □

**Lemma 6.1.** *For every  $n$ ,  $P_{\lambda_1}Y_n = 0$ .*

*Proof.* Since the urn is balanced,  $a \cdot \Delta X_n = b$  is non-random, and thus, by (4.5),

$$a \cdot Y_n := a \cdot \Delta X_{n-1} - \mathbb{E}(a \cdot \Delta X_{n-1} \mid \mathcal{F}_{n-1}) = b - b = 0. \tag{6.5}$$

The result follows by (2.12). □

Using (2.6), we can rewrite (4.16) as

$$\text{Var}(X_n) = \sum_{\lambda} \sum_{\mu} \sum_{i=1}^n F_{i,n} P_{\lambda} \mathbb{E}(Y_i Y_i') P'_{\mu} F'_{i,n}. \tag{6.6}$$

For convenience, we define

$$T_{i,n,\lambda,\mu} := F_{i,n} P_{\lambda} \mathbb{E}(Y_i Y_i') P'_{\mu} F'_{i,n}. \tag{6.7}$$

Note that Lemma 6.1 implies  $P_{\lambda_1} \mathbb{E}(Y_i Y_i') = \mathbb{E}(P_{\lambda_1} Y_i Y_i') = 0$  and thus also, by taking the transpose,  $\mathbb{E}(Y_i Y_i') P'_{\lambda_1} = 0$ . Hence  $T_{i,n,\lambda,\mu} = 0$  when  $\lambda = \lambda_1$  or  $\mu = \lambda_1$ , so these terms can be dropped and (6.6) can be written

$$\text{Var}(X_n) = \sum_{\lambda \neq \lambda_1} \sum_{\mu \neq \lambda_1} \sum_{i=1}^n T_{i,n,\lambda,\mu}. \tag{6.8}$$

We begin with a simple estimate of this sum. The same estimates are given in [7, Theorem 2.2] under similar conditions.

**Lemma 6.2.** *If  $\lambda_1 = 1$ , then, for  $n \geq 2$ ,*

$$\text{Var} X_n = \begin{cases} O(n), & \text{Re } \lambda_2 < \frac{1}{2}, \\ O(n \log^{2\nu_2+1} n), & \text{Re } \lambda_2 = \frac{1}{2}, \\ O(n^{2\text{Re } \lambda_2} \log^{2\nu_2} n), & \text{Re } \lambda_2 > \frac{1}{2}. \end{cases} \tag{6.9}$$

*In particular, if  $\lambda_2 < \lambda_1 = 1$ , then*

$$\text{Var}(X_n) = o(n^2). \tag{6.10}$$

*Proof.* It follows from (2.4) that  $\mathbb{E}(Y_n Y_n') = O(1)$ . By combining this and Lemma 5.5, we see that if  $\lambda$  and  $\mu$  are two eigenvalues, then, for  $1 \leq i \leq n$ ,

$$T_{i,n,\lambda,\mu} = F_{i,n} P_{\lambda} \mathbb{E}(Y_i Y_i') (F_{i,n} P_{\mu})' = O((n/i)^{\text{Re } \lambda + \text{Re } \mu} (1 + \log(n/i))^{\nu_{\lambda} + \nu_{\mu}}). \tag{6.11}$$

If  $\text{Re } \lambda + \text{Re } \mu \geq 1$ , we note that this implies

$$T_{i,n,\lambda,\mu} = O((n/i)^{\text{Re } \lambda + \text{Re } \mu} \log^{\nu_{\lambda} + \nu_{\mu}} n), \tag{6.12}$$

while if  $\text{Re } \lambda + \text{Re } \mu < 1$ , we choose  $\alpha$  with  $\text{Re } \lambda + \text{Re } \mu < \alpha < 1$  and note that (6.11) implies

$$T_{i,n,\lambda,\mu} = O((n/i)^{\alpha}). \tag{6.13}$$

By summing over  $i$  we obtain from (6.12) and (6.13) that

$$\sum_{i=1}^n T_{i,n,\lambda,\mu} = \begin{cases} O(n), & \text{Re } \lambda + \text{Re } \mu < 1, \\ O(n \log^{\nu_{\lambda} + \nu_{\mu} + 1} n), & \text{Re } \lambda + \text{Re } \mu = 1, \\ O(n^{\text{Re } \lambda + \text{Re } \mu} \log^{\nu_{\lambda} + \nu_{\mu}} n), & \text{Re } \lambda + \text{Re } \mu > 1. \end{cases} \tag{6.14}$$

The result (6.9) follows from (6.8) by summing (6.14) over the finitely many  $\lambda, \mu \in \sigma(A) \setminus \{\lambda_1\}$  and noting that our estimates are largest for  $\lambda = \mu = \lambda_2$ . The simpler estimate (6.10) is an immediate consequence. □

**Lemma 6.3.** *If  $\text{Re } \lambda_2 < \lambda_1 = 1$ , then, as  $n \rightarrow \infty$ ,*

$$\mathbb{E}(Y_n Y_n') \rightarrow B - \nu_1 \nu_1'. \tag{6.15}$$



Hence, for any eigenvalue  $\lambda \neq \lambda_1$ ,

$$P_\lambda \mathbb{E}(Y_n Y'_n) \rightarrow P_\lambda B. \tag{6.16}$$

*Proof.* By (4.5) and (4.4),  $Y_{n+1} = \Delta X_n - w_n^{-1} A X_n$ , with  $\mathbb{E}(Y_{n+1} | \mathcal{F}_n) = 0$  by (4.6). Hence,

$$\mathbb{E}(Y_{n+1} Y'_{n+1} | \mathcal{F}_n) = \mathbb{E}(\Delta X_n (\Delta X_n)' | \mathcal{F}_n) - w_n^{-2} A X_n (A X_n)' \tag{6.17}$$

and thus

$$\mathbb{E}(Y_{n+1} Y'_{n+1}) = \mathbb{E}(\Delta X_n (\Delta X_n)') - w_n^{-2} A \mathbb{E}(X_n X'_n) A'. \tag{6.18}$$

By the definition of the urn and (4.3),

$$\mathbb{E}(\Delta X_n (\Delta X_n)' | \mathcal{F}_n) = \sum_{j=1}^q \mathbb{P}(I_{n+1} = j | \mathcal{F}_n) \mathbb{E}(\xi_j \xi'_j) = \sum_{j=1}^q \frac{a_j X_{nj}}{w_n} \mathbb{E}(\xi_j \xi'_j),$$

and thus, using (5.1) and Theorem 3.1, and recalling (2.14), as  $n \rightarrow \infty$ ,

$$\mathbb{E}(\Delta X_n (\Delta X_n)') = \sum_{j=1}^q \frac{a_j \mathbb{E} X_{nj}}{n + w_0} \mathbb{E}(\xi_j \xi'_j) \rightarrow \sum_{j=1}^q a_j v_{1j} \mathbb{E}(\xi_j \xi'_j) = B. \tag{6.19}$$

Furthermore, by (6.10) and Theorem 3.1 again,

$$n^{-2} \mathbb{E}(X_n X'_n) = n^{-2} \text{Var}(X_n) + n^{-2} (\mathbb{E} X_n) (\mathbb{E} X_n)' \rightarrow 0 + v_1 v'_1. \tag{6.20}$$

Consequently, by (6.18), (6.19), and (6.20), and recalling that  $w_n/n \rightarrow 1$  by (5.1) and  $Av_1 = \lambda_1 v_1 = v_1$ ,

$$\mathbb{E}(Y_{n+1} Y'_{n+1}) \rightarrow B - Av_1 v'_1 A' = B - v_1 v'_1. \tag{6.21}$$

This proves (6.15), and (6.16) follows from noting that  $P_\lambda v_1 = P_\lambda P_{\lambda_1} v_1 = 0$  when  $\lambda \neq \lambda_1$ .  $\square$

*Proof of Theorem 3.2.* Let  $\lambda, \mu \in \sigma(A) \setminus \{\lambda_1\}$ , and note that, by our assumption,  $\text{Re } \lambda, \text{Re } \mu \leq \text{Re } \lambda_2 < \frac{1}{2} \lambda_1 = \frac{1}{2}$ . Write the inner sum in (6.8) as an integral:

$$\frac{1}{n} \sum_{i=1}^n T_{i,n,\lambda,\mu} = \int_0^1 T_{[xn],n,\lambda,\mu} dx. \tag{6.22}$$

For each fixed  $x \in (0, 1]$ , by Lemmas 5.7 and 6.3,

$$\begin{aligned} T_{[xn],n,\lambda,\mu} &= F_{[xn],n} P_\lambda \mathbb{E}(Y_{[xn]} Y'_{[xn]}) P'_\mu F'_{[xn],n} \\ &\rightarrow x^{-A} P_\lambda B P'_\mu x^{-A'}. \end{aligned} \tag{6.23}$$

Furthermore, choose some  $\alpha \in [0, 1)$  such that  $\text{Re } \lambda_2 < \frac{1}{2} \alpha$ . Then (6.13) applies and yields, for some  $C < \infty$ ,

$$T_{[xn],n,\lambda,\mu} \leq C(n/[xn])^\alpha \leq Cx^{-\alpha}, \tag{6.24}$$

which is integrable on  $(0,1]$ . Thus, Lebesgue’s theorem on dominated convergence applies to (6.22) and yields, by (6.23) and the change of variables  $x = e^{-s}$ ,

$$\frac{1}{n} \sum_{i=1}^n T_{i,n,\lambda,\mu} \rightarrow \int_0^1 x^{-A} P_\lambda B P'_\mu x^{-A'} dx = \int_0^\infty e^{sA} P_\lambda B P'_\mu e^{sA'} e^{-s} ds.$$

Hence, (6.8) and the definition (2.15) yield

$$\frac{1}{n} \text{Var } X_n = \frac{1}{n} \sum_{\lambda \neq \lambda_1} \sum_{\mu \neq \lambda_1} \sum_{i=1}^n T_{i,n,\lambda,\mu} \rightarrow \int_0^\infty e^{sA} \widehat{P} B \widehat{P}' e^{sA'} e^{-s} ds = \Sigma_I,$$

showing (3.3). □

*Proof of Theorem 3.3.* As in the proof of Theorem 3.2, we use (6.8) and consider the sum  $\sum_{i=1}^n T_{i,n,\lambda,\mu}$  for two eigenvalues  $\lambda, \mu \in \sigma(A) \setminus \{\lambda_1\}$ . By assumption,  $\text{Re } \lambda + \text{Re } \mu \leq 2\text{Re } \lambda_2 = 1$ , and if  $\text{Re } \lambda + \text{Re } \mu < 1$ , then  $\sum_{i=1}^n T_{i,n,\lambda,\mu} = O(n)$  by (6.14). Hence we only have to consider the case  $\text{Re } \lambda + \text{Re } \mu = 1$ , i.e.,  $\text{Re } \lambda = \text{Re } \mu = \frac{1}{2} = \text{Re } \lambda_2$ . In particular,  $\nu_\lambda, \nu_\mu \leq \nu_2$ .

In this case, as in (6.22), we transform the sum into an integral, but this time in a somewhat different way. Using the change of variables  $x = n^y = e^{y \log n}$ , we have

$$\begin{aligned} \sum_{i=1}^n T_{i,n,\lambda,\mu} &= T_{1,n,\lambda,\mu} + \int_1^n T_{\lceil x \rceil, n, \lambda, \mu} dx \\ &= T_{1,n,\lambda,\mu} + \int_0^1 T_{\lceil n^y \rceil, n, \lambda, \mu} n^y \log n dy. \end{aligned} \tag{6.25}$$

Hence, since  $T_{1,n,\lambda,\mu} = O(n \log^{2\nu_2} n)$  by (6.12),

$$(n \log^{2\nu_2+1} n)^{-1} \sum_{i=1}^n T_{i,n,\lambda,\mu} = o(1) + \int_0^1 n^{y-1} (\log n)^{-2\nu_2} T_{\lceil n^y \rceil, n, \lambda, \mu} dy. \tag{6.26}$$

Fix  $y \in (0, 1)$ . Then, by (5.21),

$$\begin{aligned} F_{\lceil n^y \rceil, n} P_\lambda &= \frac{1}{\nu_2!} \left( \frac{n}{\lceil n^y \rceil} \right)^\lambda \log^{\nu_2} \left( \frac{n}{\lceil n^y \rceil} \right) (N_\lambda^{\nu_2} P_\lambda + o(1)) \\ &= \frac{1}{\nu_2!} n^{(1-y)\lambda} ((1-y) \log n)^{\nu_2} (N_\lambda^{\nu_2} P_\lambda + o(1)), \end{aligned} \tag{6.27}$$

and similarly for  $\mu$ .

Recall the assumption  $\text{Re } \lambda + \text{Re } \mu = 1$ , and let  $\tau := \text{Im } \lambda + \text{Im } \mu$ , so  $\lambda + \mu = 1 + i\tau$ . Then, by (6.7), (6.27), and (6.16),

$$\begin{aligned} &n^{y-1} (\log n)^{-2\nu_2} T_{\lceil n^y \rceil, n, \lambda, \mu} \\ &= \frac{1}{(\nu_2!)^2} n^{i(1-y)\tau} (1-y)^{2\nu_2} N_\lambda^{\nu_2} P_\lambda B (N_\mu^{\nu_2} P_\mu)' + o(1). \end{aligned} \tag{6.28}$$

Moreover, by (6.12), uniformly for  $y \in (0, 1]$  and  $n \geq 2$ ,

$$n^{y-1} (\log n)^{-2\nu_2} T_{\lceil n^y \rceil, n, \lambda, \mu} = O((n/\lceil n^y \rceil) n^{y-1}) = O(1). \tag{6.29}$$

Hence the error term  $o(1)$  in (6.28) is also uniformly bounded, and we can apply dominated convergence to the integral of it, yielding

$$\int_0^1 n^{y-1}(\log n)^{-2\nu_2} T_{\lceil n^y \rceil, n, \lambda, \mu} dy = \frac{1}{(\nu_2!)^2} \int_0^1 n^{i(1-y)\tau} (1-y)^{2\nu_2} dy \cdot N_\lambda^{\nu_2} P_\lambda B(N_\mu^{\nu_2} P_\mu)' + o(1). \tag{6.30}$$

In the case  $\tau = 0$ , i.e.,  $\mu = \bar{\lambda}$ , the integral on the right-hand side of (6.30) is  $\int_0^1 (1-y)^{2\nu_2} dy = (2\nu_2 + 1)^{-1}$ . Furthermore, in this case,  $P_\mu = P_{\bar{\lambda}} = \bar{P}_\lambda$  and thus  $P'_\mu = P^*_\lambda$ , and similarly  $N'_\mu = N^*_\lambda$ . Hence, (6.31) yields

$$\int_0^1 n^{y-1}(\log n)^{-2\nu_2} T_{\lceil n^y \rceil, n, \lambda, \bar{\lambda}} dy = \frac{1}{(2\nu_2 + 1)(\nu_2!)^2} N_\lambda^{\nu_2} P_\lambda B P^*_\lambda (N^*_\lambda)^{\nu_2} + o(1). \tag{6.31}$$

On the other hand, if  $\tau \neq 0$ , then, with  $u = 1 - y$ ,

$$\int_0^1 n^{i(1-y)\tau} (1-y)^{2\nu_2} dy = \int_0^1 e^{i(\tau \log n)u} u^{2\nu_2} du \rightarrow 0 \tag{6.32}$$

as  $n \rightarrow \infty$ , and thus  $|\tau \log n| \rightarrow \infty$ , by an integration by parts (or by the Riemann–Lebesgue lemma). Hence, when  $\mu \neq \bar{\lambda}$ , (6.30) yields

$$\int_0^1 n^{y-1}(\log n)^{-2\nu_2} T_{\lceil n^y \rceil, n, \lambda, \mu} dy = o(1). \tag{6.33}$$

We saw in the beginning of the proof that we can ignore the terms in (6.8) with  $\text{Re } \lambda < \frac{1}{2}$  or  $\text{Re } \mu < \frac{1}{2}$ , and by (6.26) and (6.33), we can also ignore the case  $\text{Re } \lambda = \text{Re } \mu = \frac{1}{2}$  but  $\mu \neq \bar{\lambda}$ . Hence only the case  $\mu = \bar{\lambda}$  with  $\text{Re } \lambda = \frac{1}{2}$  remains in (6.8), and the result follows by (6.26) and (6.33). □

*Proof of Theorem 3.4.* By (6.10),

$$\mathbb{E} \|X_n/n - \mathbb{E} X_n/n\|^2 = n^{-2} \mathbb{E} \|X_n - \mathbb{E} X_n\|^2 = \sum_{i=1}^q n^{-2} \text{Var}(X_{ni}) \rightarrow 0,$$

and  $\mathbb{E} X_n/n \rightarrow \nu_1$  by Theorem 3.1. Hence,  $\mathbb{E} \|X_n/n - \nu_1\|^2 \rightarrow 0$ , which is the claimed convergence in  $L^2$ .

Moreover, if we fix  $\varepsilon \in (0, \frac{1}{2})$  such that  $\text{Re } \lambda_2 < 1 - \varepsilon$ , then the same argument shows, using (6.9) and (6.1)–(6.4), that, more precisely,

$$\mathbb{E} \|X_n - (n + w_0)\nu_1\|^2 = O(n^{2-2\varepsilon}). \tag{6.34}$$

Let  $N \geq 1$ . By (4.11) and the definition (4.10), for any  $n \leq N$ ,

$$F_{n,N} X_n = F_{0,N} X_0 + \sum_{\ell=1}^n F_{\ell,N} Y_\ell. \tag{6.35}$$

Moreover, by (4.6),  $Y_n$  is a martingale difference sequence, and thus so is  $F_{n,N} Y_n$ , for  $n \leq N$ . Hence, (6.35) shows that  $F_{n,N} X_n$ ,  $n \leq N$ , is a martingale, and thus

$$F_{n,N} X_n = \mathbb{E}(X_N | \mathcal{F}_n), \quad n \leq N. \tag{6.36}$$

By Lemma 5.6,

$$F_{n,N}v_1 = F_{n,N}P_{\lambda_1}v_1 = \frac{N + w_0}{n + w_0}v_1,$$

and thus (6.36) implies

$$F_{n,N}(X_n - (n + w_0)v_1) = \mathbb{E}(X_N - (N + w_0)v_1 \mid \mathcal{F}_n), \quad n \leq N. \tag{6.37}$$

Hence, by Doob’s inequality (applied to each coordinate) and (6.34),

$$\mathbb{E} \sup_{n \leq N} \|F_{n,N}(X_n - (n + w_0)v_1)\|^2 \leq 4\mathbb{E}\|X_N - (N + w_0)v_1\|^2 = O(N^{2-2\varepsilon}). \tag{6.38}$$

It follows, using Lemma 5.8, that if  $N \geq 2i_0$ , then

$$\mathbb{E} \sup_{N/2 \leq n \leq N} \|(X_n - (n + w_0)v_1)/n\|^2 = O(N^{-2\varepsilon}). \tag{6.39}$$

This holds trivially for smaller  $N$  as well, since each  $X_n \in L^2$  and thus the left-hand side of (6.41) is finite for each  $N$ . Consequently, taking  $N = 2^k$  and summing,

$$\mathbb{E} \sum_{k=1}^{\infty} \sup_{2^{k-1} \leq n \leq 2^k} \|(X_n - (n + w_0)v_1)/n\|^2 < \infty. \tag{6.40}$$

Consequently,  $\|(X_n - (n + w_0)v_1)/n\| \rightarrow 0$  a.s., and thus  $X_n/n \xrightarrow{\text{a.s.}} v_1$ . □

*Proof of Theorem 3.5.* If  $\text{Var}(u \cdot X_n) = 0$  for every  $n$ , then  $u' \Sigma u = 0$  by (3.4).

For the converse, assume that  $u' \Sigma u = 0$ . Then, by (3.3) and (2.15),

$$0 = u' \Sigma_I u = \int_0^{\infty} u' \widehat{P} e^{sA} B e^{sA'} \widehat{P}' u e^{-\lambda_1 s} ds. \tag{6.41}$$

The integrand is a continuous function of  $s \geq 0$ , and non-negative since  $B$  is non-negative definite by (2.14). Hence, the integrand vanishes for every  $s \geq 0$ . In particular, taking  $s = 0$  we obtain, using (2.14) again,

$$\begin{aligned} 0 &= u' \widehat{P} B \widehat{P}' u = \sum_{i=1}^q a_i v_{1i} u' \widehat{P} \mathbb{E}(\xi_i \xi_i') \widehat{P}' u \\ &= \sum_{i=1}^q a_i v_{1i} \mathbb{E}(u' \widehat{P} \xi_i (u' \widehat{P} \xi_i)') = \sum_{i=1}^q a_i v_{1i} \mathbb{E}(u' \widehat{P} \xi_i)^2, \end{aligned} \tag{6.42}$$

noting that  $u' \widehat{P} \xi_i$  is a scalar. Each term is non-negative, and thus each term is 0. If  $i$  is such that  $a_i > 0$ , then it follows from the assumption that  $A$  is irreducible that  $v_{1i} > 0$ , and hence (6.42) yields  $\mathbb{E}(u' \widehat{P} \xi_i)^2 = 0$  and thus  $u' \widehat{P} \xi_i = 0$  a.s. Furthermore, since the urn is balanced, by (2.12),

$$P_{\lambda_1} \xi_i = (a \cdot \xi_i)v_1 = b v_1. \tag{6.43}$$

Hence, for every  $i$  with  $a_i > 0$ ,

$$u \cdot \xi_i = u \cdot (\widehat{P} + P_{\lambda_1}) \xi_i = 0 + u \cdot (b v_1) = b u \cdot v_1. \tag{6.44}$$

This is independent of  $i$ , and thus, for every  $n$ , a.s.,

$$u \cdot \Delta X_n = bu \cdot v_1. \tag{6.45}$$

Consequently, a.s.,

$$u \cdot X_n = u \cdot X_0 + nbu \cdot v_1 \tag{6.46}$$

and thus  $u \cdot X_n$  is deterministic. □

### 7. Examples

Pólya urns have been used for a long time in various applications, for example to study fringe structures in various random trees; see e.g. [5], [2], and [11]. Some recent examples are given in [15], where, in particular, the number of two-protected nodes in a random  $m$ -ary search tree is studied for  $m = 2$  and  $m = 3$  using suitable Pólya urns with 5 and 19 types, respectively, and it is shown that if this number is denoted by  $Y_n$  ( $m = 2$ ) or  $Z_n$  ( $m = 3$ ) for a search tree with  $n$  keys, then

$$\frac{Y_n - \frac{11}{30}n}{\sqrt{n}} \xrightarrow{d} N\left(0, \frac{29}{225}\right), \tag{7.1}$$

$$\frac{Z_n - \frac{57}{700}n}{\sqrt{n}} \xrightarrow{d} N\left(0, \frac{1692302314867}{43692253605000}\right). \tag{7.2}$$

(The binary case (7.1) was shown earlier by Mahmoud and Ward [25] using other methods.) The urns are strictly small; in both cases  $\lambda_1 = 1$  and  $\lambda_2 = 0$ , with  $v_2 = 0$ , and Theorems 3.1 and 3.2 yield, using the calculations in [15] (see Remark 3.1),

$$\mathbb{E}Z_n = \frac{57}{700}n + O(1), \tag{7.3}$$

$$\text{Var } Z_n = \frac{1692302314867}{43692253605000}n + o(n), \tag{7.4}$$

together with corresponding results for  $Y_n$ . (The results for  $Y_n$  were previously shown in Mahmoud and Ward [25], where exact formulas for the mean and variance of  $Y_n$  were given.)

Furthermore, [15] also studies the numbers of leaves and one-protected nodes in a random  $m$ -ary search tree using a similar but simpler urn. (For  $m = 2$  this was done already by Devroye [11].) For  $2 \leq m \leq 26$ , this is a strictly small urn, and again the results in Section 3 yield asymptotics of mean and variance.

See [16] for further similar examples.

**Remark 7.1.** As mentioned above, the urn used to show (7.1) has five types, corresponding to five different small trees. Drawing a ball corresponds to adding a node to a (randomly chosen) gap in the corresponding tree; this may cause the tree to break up into several smaller trees. The five types have 4, 3, 2, 1, and 0 gaps, respectively, and these numbers are their activities. Moreover, for type 2, the gaps are not equivalent, which makes the replacement for this type random. (We have  $\xi_2 = (1, -1, 0, 0, 0)$  with probability  $1/3$  and  $\xi_2 = (0, 0, 0, 1, 0)$  with probability  $2/3$ ; see [15].)

A different, essentially equivalent, approach is to instead consider as types the different gaps in the different trees; this yields five new types that we denote by 1, 2A, 2B, 3, 4. The

transition from the old urn to the new is a simple linear transformation: each old ball of type 1 is replaced by four new balls of type 1, which we write as  $1 \rightarrow 4 \cdot 1$ , and similarly  $2 \rightarrow 2A + 2 \cdot 2B$ ,  $3 \rightarrow 2 \cdot 3$ ,  $4 \rightarrow 4$ , while balls of type 5 (which has activity 0) are ignored. This yields a new Pólya urn, where all types have activity 1. In the new urn, all replacements are deterministic, which sometimes is an advantage, but on the other hand, replacements now may involve subtractions. For example, in the original urn,  $\xi_1 = (-1, 1, 1, 0, 0)$ , meaning that if we draw a ball of type 1, it is discarded and replaced by one of type 2 and one of type 3. In the new urn, this translates to  $\xi_1 = (-4, 1, 2, 2, 0)$ , meaning that we remove the drawn ball together with three others of type 1, and then add  $2A + 2 \cdot 2B + 2 \cdot 3$ . Even worse,  $\xi_2 = (4, -1, -2, 0, 0)$ , meaning that if we draw a ball of type 2A, we remove it together with two balls of type 2B, and add four balls of type 1. Nevertheless, by the construction, the urn is obviously tenable in the sense of the present paper. This urn, with the gaps as types, thus is an example of a tenable urn with subtractions that occur naturally in an application.

The Pólya urn for the ternary search tree with 19 types in [15] can similarly be translated into an urn (with 29 types) using gaps as types, again with deterministic replacements, but sometimes subtractions.

See also [15], where the transition to the corresponding urn with gaps was used for the simpler urn used to study leaves; in that case there are no subtractions.

### 8. Further comments

The decomposition (4.11) and its consequence (4.16) explain some of the differences between the small and large urns described in the introduction. Suppose again for convenience that  $\lambda_1 = 1$ . Then the term  $F_{\ell,n}Y_\ell$  in (4.11), which is the (direct and indirect) contribution from the  $\ell$ th draw, has a variance roughly (ignoring logarithmic factors when  $\nu_2 > 0$ ) of the order  $(n/\ell)^{2\text{Re } \lambda_2}$ ; see Lemma 6.2 and its proof. For a large urn, this decreases rapidly with  $\ell$ , and  $\sum_\ell \ell^{-2\text{Re } \lambda_2}$  converges; thus the variance is dominated by the contribution from the first draws. This strong long-term dependency leads to the a.s. limit results, and to the dependency of the limit on the initial state  $X_0$ .

On the other hand, for a strictly small urn, the sum of the variances is of the order  $n$ , but each term is  $o(n)$  and is negligible, which explains why the first draws, and the initial state, do not affect the limit distribution. In fact, for a component  $X_{n,i}$  with asymptotic variance  $(\Sigma)_{ii} > 0$ , we see that for any  $\varepsilon > 0$ , all but a fraction  $\varepsilon$  of  $\text{Var } X_{n,i}$  is explained by the draws with numbers in  $[\delta n, n]$ , for some  $\delta = \delta(\varepsilon) > 0$ . The long-term dependency is thus weak in this case.

The remaining case, a small urn with  $\text{Re } \lambda_2 = 1/2$ , is similar to the strictly small case, but the long-term dependency is somewhat stronger. If for simplicity we assume  $\nu_2 = 0$ , then the contribution of the  $\ell$ th draw to  $\text{Var}(X_n)$  is of the order  $n/\ell$ , giving a total variance of order  $n \log n$ . Again, the first draws and the initial state do not affect the limit distribution, but in order to explain all but a fraction  $\varepsilon$  of the variance, we have to use the draws in  $[n^\delta, n]$ , for some small  $\delta > 0$ . (Cf. the functional limit theorem [19, Theorem 3.31] with different time scales for strictly small and non-strictly small urns.)

Cf. also [19, Remark 4.3], where a similar argument is made using the corresponding continuous-time branching process.

### Appendix A. The largest eigenvalue

We have seen in (2.8) that for a balanced urn,  $b$  is an eigenvalue of  $A$ , with a non-negative left eigenvector  $a$ .

In typical applications,  $b$  is the largest eigenvalue  $\lambda_1$ . Before proceeding, let us note that this is not always true.

**Example A.1.** As a counterexample, consider an urn with three colours, with activities 1, and the (deterministic) replacements  $\xi_1 = (1, 2, 0)$ ,  $\xi_2 = (2, 1, 0)$ ,  $\xi_3 = (-1, 0, 4)$ . The urn is balanced, with  $b = 3$ , and if we start with  $X_0 = (1, 0, 0)$ , the urn is tenable. Nevertheless, the largest eigenvalue  $\lambda_1 = 4 > b$ . Of course, the reason is that the urn never will contain any ball of colour 3, so this ought to be treated as an urn with just colours 1 and 2. (If there is any ball of colour 3, the urn is not tenable.)

Example A.1 is obviously a silly counterexample, but it shows that we need some extra assumption to exclude such trivialities. We have the following result, which shows that if we only use colours that actually can occur, then  $\lambda_1 = b$  holds (or, at least, may be assumed) and  $v_1 = 0$ .

**Lemma A.1.** *If the Pólya urn is tenable and balanced, and moreover every colour has a non-zero probability of ever appearing in the urn, then  $\text{Re } \lambda \leq b$  for every  $\lambda \in \sigma(A)$ , and, furthermore, if  $\text{Re } \lambda = b$  then  $v_\lambda = 0$ . We may thus assume  $\lambda_1 = b$ .*

*Proof.* As in Section 5, we may and shall assume that  $b = 1$ .  
 Suppose that  $\lambda \in \sigma(A)$ . By (4.12) and Lemma 5.4,

$$\begin{aligned} P_\lambda \mathbb{E}X_n &= P_\lambda F_{0,n} X_0 = F_{0,n} P_\lambda X_0 \\ &= n^\lambda \log^{v_\lambda} n \frac{\Gamma(w_0)}{v_\lambda! \Gamma(w_0 + \lambda)} N_\lambda^{v_\lambda} P_\lambda X_0 + o(n^{\text{Re } \lambda} \log^{v_\lambda} n). \end{aligned} \tag{A.1}$$

On the other hand, by our assumption (2.4),  $\mathbb{E}\|\xi_i\| < \infty$  for each  $i$ , and thus  $\mathbb{E}\|\Delta X_n\| \leq \max_i \mathbb{E}\|\xi_i\| < \infty$  and therefore  $\|\mathbb{E}X_n\| \leq \mathbb{E}\|X_n\| = O(n)$ . Hence,

$$P_\lambda \mathbb{E}X_n = O(n). \tag{A.2}$$

Suppose now that either  $\text{Re } \lambda > 1$ , or  $\text{Re } \lambda = 1$  and  $v_\lambda > 0$ . Then (A.1) and (A.2) yield the desired contradiction unless

$$N_\lambda^{v_\lambda} P_\lambda X_0 = 0. \tag{A.3}$$

Moreover, if we run the urn for  $k$  steps and regard  $X_k$  as a new starting position (conditioning on  $X_k$ ), then the resulting urn is a.s. tenable; hence the argument just given shows that

$$N_\lambda^{v_\lambda} P_\lambda X_k = 0 \tag{A.4}$$

a.s. for every  $k \geq 0$ . Hence, also  $N_\lambda^{v_\lambda} P_\lambda \Delta X_k = 0$  a.s. If  $j$  is any colour with  $a_j > 0$ , then, by assumption, there exists a  $k$  such that  $\mathbb{P}(X_{k,j} > 0) > 0$ , and thus with positive probability  $I_{k+1} = j$  and then  $\Delta X_k$  is a copy of  $\xi_j$ . Consequently, if  $a_j > 0$  then

$$N_\lambda^{v_\lambda} P_\lambda \xi_j = 0 \tag{A.5}$$

a.s., and thus

$$N_\lambda^{v_\lambda} P_\lambda \mathbb{E}\xi_j = 0. \tag{A.6}$$

In other words, for every  $j$ ,

$$N_\lambda^{v_\lambda} P_\lambda (a_j \mathbb{E}\xi_j) = 0. \tag{A.7}$$



However, by (2.5), the  $j$ th column of  $A$  is  $a_j \mathbb{E} \xi_j$ . Consequently, (A.7) is equivalent to  $N_\lambda^{v_\lambda} P_\lambda A = 0$ . Since  $P_\lambda A = \lambda P_\lambda + N_\lambda$  by (2.7), and  $N_\lambda^{v_\lambda+1} = 0$ , this yields

$$0 = N_\lambda^{v_\lambda} P_\lambda A = \lambda N_\lambda^{v_\lambda} P_\lambda + N_\lambda^{v_\lambda+1} = \lambda N_\lambda^{v_\lambda} \tag{A.8}$$

and thus  $\lambda = 0$ , which contradicts the assumption on  $\lambda$ .

Consequently,  $\operatorname{Re} \lambda \leq 1 = b$  for every  $\lambda \in \sigma(A)$ , and since  $b \in \sigma(A)$ ,  $\operatorname{Re} \lambda_1 = \max_{\lambda \in \sigma(A)} \operatorname{Re} \lambda = b$ . We have not ruled out the possibility that there are other eigenvalues  $\lambda$  with  $\operatorname{Re} \lambda = b$  (see Remark A.2 below), but even if this should happen, we have shown that they all have  $v_\lambda = 0$ , so we are free to choose  $\lambda_1 = b$ .  $\square$

**Remark A.1.** As noted in Remark 2.2, the eigenvalue  $\lambda_1 = b$  may be multiple. Lemma A.1 shows that  $v_1 := v_{\lambda_1} = 0$  also in this case.

**Remark A.2.** Lemma A.1 is not completely satisfactory since it does not rule out the possibility that besides  $b$ , there is also some complex eigenvalue  $\lambda = b + it$  with  $t \neq 0$ . We do not believe that this is possible, but we do not know a proof for a general tenable urn under our assumptions.

### Appendix B. A note on (2.15)

In the balanced case, by (2.12) and (2.3), a.s.,

$$P_{\lambda_1} \xi_i = (a \cdot \xi_i) v_1 = b v_1 = \lambda_1 v_1, \tag{B.1}$$

and thus, by (2.14) and (2.5),

$$\begin{aligned} P_{\lambda_1} B &= \sum_{i=1}^q a_i v_1 i \mathbb{E}(P_{\lambda_1} \xi_i \xi_i') = \sum_{i=1}^q a_i v_1 i \lambda_1 v_1 \mathbb{E}(\xi_i') = \lambda_1 v_1 \left( \sum_{i=1}^q a_i v_1 i \mathbb{E} \xi_{ij} \right)' \\ &= \lambda_1 v_1 \left( \sum_{i=1}^q (A)_{ji} v_1 i \right)' = \lambda_1 v_1 (A v_1)' = \lambda_1^2 v_1 v_1'. \end{aligned} \tag{B.2}$$

Since  $B$  is symmetric, also  $B P_{\lambda_1}' = (P_{\lambda_1} B)' = \lambda_1^2 v_1 v_1'$ , and thus

$$P_{\lambda_1} B P_{\lambda_1}' = P_{\lambda_1} B = B P_{\lambda_1}' = \lambda_1^2 v_1 v_1' \tag{B.3}$$

and, as a simple consequence, still in the balanced case,

$$\widehat{P} B \widehat{P}' = \widehat{P} B = B \widehat{P}' = B - \lambda_1^2 v_1 v_1'. \tag{B.4}$$

Cf. Lemma 6.3 (where  $\lambda_1 = 1$ ).

Hence, in the balanced case, we can omit either  $\widehat{P}$  or  $\widehat{P}'$  in (2.15). (This was noted empirically by Axel Heimbürger and Cecilia Holmgren, in personal communication.) However, by (B.4), we cannot omit both, nor even move both outside the integral, because  $e^{sA} v_1 = e^{\lambda_1 s} v_1$  and thus

$$\int_0^\infty e^{sA} v_1 v_1' e^{sA'} e^{-\lambda_1 s} ds = \int_0^\infty e^{\lambda_1 s} v_1 v_1' ds, \tag{B.5}$$

which diverges.

### Acknowledgements

I thank the anonymous referees for insightful and helpful comments. This work was supported by the Knut and Alice Wallenberg Foundation.

### References

- [1] ALDOUS, D. J. AND EAGLESON, G. K. (1978). On mixing and stability of limit theorems. *Ann. Prob.* **6**, 325–331.
- [2] ALDOUS, D. J., FLANNERY, B. AND PALACIOS, J. L. (1988). Two applications of urn processes—the fringe analysis of search trees and the simulation of quasi-stationary distributions of Markov chains. *Prob. Eng. Inf. Sci.* **2**, 293–307.
- [3] 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.
- [4] ATHREYA, K. B. AND NEY, P. E. (1972). *Branching Processes*. Springer, Berlin.
- [5] 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.
- [6] BAI, Z.-D. AND HU, F. (1999). Asymptotic theorems for urn models with nonhomogeneous generating matrices. *Stoch. Process. Appl.* **80**, 87–101.
- [7] BAI, Z.-D. AND HU, F. (2005). Asymptotics in randomized urn models. *Ann. Appl. Prob.* **15**, 914–940.
- [8] BAI, Z.-D., HU, F. AND ROSENBERGER, W. F. (2002). Asymptotic properties of adaptive designs for clinical trials with delayed response. *Ann. Statist.* **30**, 122–139.
- [9] BERNSTEIN, S. N. (1940). Nouvelles applications des grandeurs aléatoires presque indépendantes. *Izv. Akad. Nauk SSSR Ser. Mat.* **4**, 137–150 (in Russian).
- [10] BERNSTEIN, S. N. (1940). Sur un problème du schéma des urnes à composition variable. *C. R. Acad. Sci. URSS* **28**, 5–7.
- [11] DEVROYE, L. (1991). Limit laws for local counters in random binary search trees. *Random Structures Algorithms* **2**, 303–315.
- [12] DUNFORD, N. AND SCHWARTZ, J. T. (1958). *Linear Operators. Part I: General Theory*. Interscience Publishers, Inc., New York.
- [13] EGGENBERGER, F. AND PÓLYA, G. (1923). Über die Statistik verketteter Vorgänge. *Z. Angew. Math. Mech.* **3**, 279–289.
- [14] FLAJOLET, P., GABARRÓ, J. AND PEKARI, H. (2005). Analytic urns. *Ann. Prob.* **33**, 1200–1233.
- [15] HOLMGREN, C. AND JANSON, S. (2015). Asymptotic distribution of two-protected nodes in ternary search trees. *Electron. J. Prob.* **20**, 1–20.
- [16] HOLMGREN, C., JANSON, S. AND ŠILEIKIS, M. (2017). Multivariate normal limit laws for the numbers of fringe subtrees in  $m$ -ary search trees and preferential attachment trees. *Electron. J. Combin.* **24**, Paper 2.51, 49 pp.
- [17] HU, F. AND ROSENBERGER, W. F. (2006). *The Theory of Response-Adaptive Randomization in Clinical Trials*. Wiley-Interscience, Hoboken, NJ.
- [18] HU, F. AND ZHANG, L.-X. (2004). Asymptotic normality of urn models for clinical trials with delayed response. *Bernoulli* **10**, 447–463.
- [19] JANSON, S. (2004). Functional limit theorems for multitype branching processes and generalized Pólya urns. *Stoch. Process. Appl.* **110**, 177–245.
- [20] JANSON, S. (2005). Limit theorems for triangular urn schemes. *Prob. Theory Relat. Fields* **134**, 417–452.
- [21] JANSON, S. AND POUYANNE, N. (2018). Moment convergence of balanced Pólya processes. *Electron. J. Prob.* **23**, paper no. 34, 13 pp.
- [22] JIŘINA, M. (1958). Stochastic branching processes with continuous state space. *Czechoslovak Math. J.* **8**, 292–313.
- [23] JOHNSON, N. L. AND KOTZ, S. (1977). *Urn Models and Their Application*. Wiley, New York.
- [24] MAHMOUD, H. M. (2009). *Pólya Urn Models*. CRC Press, Boca Raton, FL.
- [25] MAHMOUD, H. M. AND WARD, M. D. (2012). Asymptotic distribution of two-protected nodes in random binary search trees. *Appl. Math. Lett.* **25**, 2218–2222.
- [26] MARKOV, A. A. (1917). Sur quelques formules limites du calcul des probabilités. *Bull. Acad. Imp. Sci.* **11**, 177–186 (in Russian).
- [27] OLVER, F. W. J., LOZIER, D. W., BOISVERT, R. F. AND CLARK, C. W. (eds) (2010). *NIST Handbook of Mathematical Functions*. Cambridge University Press. Also available as *NIST Digital Library of Mathematical Functions*, <http://dlmf.nist.gov>.
- [28] PÓLYA, G. (1930). Sur quelques points de la théorie des probabilités. *Ann. Inst. H. Poincaré* **1**, 117–161.

- [29] POUYANNE, N. (2008). An algebraic approach to Pólya processes. *Ann. Inst. H. Poincaré Prob. Statist.* **44**, 293–323.
- [30] SAVKEVITCH, V. (1940). Sur le schéma des urnes à composition variable. *C. R. Acad. Sci. URSS* **28**, 8–12.
- [31] ZHANG, L.-X., HU, F. AND CHEUNG, S. H. (2006). Asymptotic theorems of sequential estimation-adjusted urn models. *Ann. Appl. Prob.* **16**, 340–369.