MOMENTS OF BALANCED PÓLYA URNS

SVANTE JANSON

Abstract. We give bounds for (central) moments for balanced Pólya urns under very general conditions. In some cases, these bounds imply that moment convergence holds in earlier known results on asymptotic distribution. The results overlap with previously known results, but are here given more generally and with a simpler proof.

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 (possibly random) set of balls that depends on the colour of the drawn balls. This is repeated *ad infinitum*, and we study the asymptotic composition of the urn. For details, and the assumptions used in the present paper, see Section 2.1.

Asymptotic results for Pólya urns, including asymptotic distributions (after suitable normalization), have been proved by many authors under various conditions and in varying generality, beginning with the pioneering papers by Markov [17] and Eggenberger and Pólya [9]; for the history of Pólya urns, see e.g. Mahmoud [16]; see also the references in [11] and [14].

It is well-known that the asymptotic behaviour of a Pólya urn depends on the eigenvalues of the *intensity matrix* of the urn defined in (2.4) below, and in particular on the two largest (in real part) eigenvalues λ_1 and λ_2. In particular, if the urn satisfies some irreducibility condition (and some technical conditions) there is a dichotomy (or trichotomy if the critical case $\operatorname{Re}\lambda_2 = \frac{1}{2}\lambda_1$ is considered separately):

(i) If $\operatorname{Re}\lambda_2 \leqslant \frac{1}{2}\lambda_1$ (a *small urn*), then the number of balls of a given colour is asymptotically normal.

(ii) If $\operatorname{Re}\lambda_2 > \frac{1}{2}\lambda_1$ (a *large urn*), then this is not true: then there are 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 not the full sequence.

See for example [11, Theorems 3.22–3.24] for general results of this type. Pouyanne [18] showed more precise results for large urns (not necessarily irreducible) that are *balanced*, see Section 2.1. As another example, for urns that are triangular (and thus not irreducible), the asymptotic distribution still depends in important ways on the eigenvalues of A, but in a more complicated way, see [14]. (Note that we see again a dichotomy between

Date: 20 May, 2025.

2010 *Mathematics Subject Classification.* 60C05; 60F25.

Partly supported by the Knut and Alice Wallenberg Foundation.

small and large urns for the triangular urns in [12, Theorem 1.3(i)–(iii)] but not in [12, Theorem 1.3(iv)–(v)].)

In the present paper we give no results on asymptotic distributions; instead we consider as a complement the question whether moment convergence holds in such results on asymptotic distributions. This too is not a new subject. For example, for balanced small urns with 2 colours, Bernstein [5, 6] showed asymptotic normality in the small urn case and gave results on mean and variance; Bagchi and Pal [2] (independently, but 45 years later) gave another proof of asymptotic normality using the method of moments and thus proving moment convergence as part of the proof. Other examples are Bai and Hu [3, 4], who consider a version of Pólya urns allowing time-dependent replacements and show for small urns both asymptotic normality and (implicitly) asymptotic results for mean and variance. The results by Pouyanne [18] for balanced large urns include convergence in L^p for any p, and thus moment convergence. Janson and Pouyanne [15] proved moment convergence for irreducible small urns; the method there combined the known result on asymptotic normality in this case, and moment estimates by the method of [18] leading to uniform integrability. Janson [13] gave asymptotics for mean and variance of balanced small urns (consistent with known results on asymptotic normality in the irrdeucible case, but more general).

In the present paper (as in many of the references above) we consider only *balanced* urns, but otherwise our conditions are very general. (For example, we allow random replacements.) The main purpose of the paper is to prove general *bounds* for moments of an arbitrary fixed order (see Section 3); these bounds imply uniform integrability and can thus in many cases be combined with known results on convergence in distribution to show that moment convergence hold in the latter results (see for example Theorem 3.2). Our results are to a large extent not new, but as far as we know stated in greater generality than earlier; moreover, the method of proof seems simpler that the ones used earlier. For earlier related results, as said above, Pouyanne [18] showed moment convergence for balanced large urns; Janson and Pouyanne [15] showed moment bounds (and thus convergence) for balanced small irreducible urns; Janson [13] treated first and second moments for balanced (and often small) urns – the method used here is a further development and simplification of the method in [13].

We do not make any assumptions on the structure of the urn in this paper (except being balanced). However, the results are particularly adapted to the irreducible case, where the bounds that we obtain match the normalizations in the known results on asymptotic normality, thus showing that these results hold with moment convergence. The results below apply also to, for example, balanced triangular urns, but in that case the results are in general less sharp and there is often a gap between the general bounds proved here and the moment convergence proved for such urns in [14, Theorem 12.5]. For example, for a balanced triangular urn with 2 colours, see [14, Example 14.4 Case 4] and [12, Theorem 1.3(v)], the bounds below are sharp if the urn is large but not if it is small.

Our results for balanced urns are very general, but that leaves one obvious open problem:

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 unbalanced urns.

In fact, it seems that only essentially trivial examples are known of an unbalanced urn where (after suitable normalization) convergence in distribution holds with all moments. In the negative direction, [14, Example 14.2] gives an example (an unbalanced diagonal urn) with convergence in distribution to a limit with infinite mean; hence moment convergence does not hold. However, this counterexample seems rather special and is maybe not typical.

Section 2 gives definitions and some other preliminaries. Section 3 contains the statements of the main results, which are proved in Sections 4–6. Appendix A gives some further, more technical, results, and Appendix B a proof of a simple lemma for which we do not know a reference.

Acknowledgement. This paper owes much to an anonymous referee of my earlier paper [13] dealing with first and second moments. The referee suggested an alternative method of proof of results there, using Lemma A.2 below, and also said that this could be extended to prove convergence of higher moments. I am very grateful for that suggestion, which (although not used in [13]) I have developed and further simplified, leading to the present paper.

2. Preliminaries and notation

2.1. **Definition and assumptions.** A (generalized) Pólya urn process is defined as a discrete-time Markov process of the following type:

(PU1) The state of the urn at time n is given by the vector $X_n = (X_{n1}, \dots, X_{nq}) \in [0,\infty)^q$ for some given integer $q \geqslant 2$, where X_{ni} is interpreted as the number of balls of colour i in the urn; thus balls can have the q colours (types) $1, \dots, q$. The urn starts with a given vector X_0.

(PU2) Each colour i has a given *activity* (or weight) $a_i \geqslant 0$, and a (generally random) *replacement vector* $\xi_i = (\xi_{i1}, \dots, \xi_{iq})$. At each time $n+1 \geqslant 1$, 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 colour i, it is replaced together with ΔX_{nj} balls of colour j, $j = 1, \dots, q$, where the random vector $\Delta X_n = (\Delta X_{n1}, \dots, \Delta X_{nq})$ has the same distribution as ξ_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_{nj} and ξ_{ij} are integers, but that is not necessary; as is well-known, the Pólya urn process is well-defined also for *real* X_{ni} and ξ_{ij}, with probabilities for the different replacements still given by (2.1); the "number of balls" X_{ni} may thus be any nonnegative 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 ξ_{ij} are thus in general 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 ξ_i are such that, for every $n \geqslant 0$, a.s.

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

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*.

Note that we allow some activities $a_i = 0$ (as long as (2.2) holds); this means that balls of colour i never are drawn. (This is useful in some applications, see e.g. [11].)

For simplicity, we assume throughout this paper also:

(PU3) The initial state X_0 is nonrandom.

Remark 2.1. The results below can be extended to random X_0 by conditioning on X_0, but we have not checked exactly what conditions are needed, and we leave this to the reader.

We will in the present paper (as in [13]) only consider balanced Pólya urns, defined as follows:

(PU4) 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 (nonrandom and not depending on the colour of the drawn ball).

The balance condition (PU4) together with (PU3) imply that the denominator in (2.1) (i.e., the total activity in the urn) is deterministic for each n, see (4.2)–(4.3). This is a significant simplification, assumed in many papers on Pólya urns, and luckily satisfied in many applications.

2.2. **Some notation.** We regard all vectors as column vectors. We use standard notations for (real or complex) vectors and matrices (of sizes q and $q \times q$, respectively), in particular $'$ for transpose; we also use $\cdot$ for the bilinear scalar product defined by $u \cdot v = u'v$ for any vectors $u, v \in \mathbb{C}^q$. (This is the standard scalar product for $u, v \in \mathbb{R}^q$, but note the abscence of conjugation in general.) We denote the standard Euclidean norm for vectors by $|\cdot|$, and denote the operator norm for matrices by $\|\cdot\|$.

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

We let $[q] := \{1, \dots, q\}$, the (finite) set of colours.

For a random variable (or vector) X, we denote the usual L^p-norm by $\|X\|_p := (\mathbb{E}\,|X|^p)^{1/p}$, $p \geqslant 1$.

We let C denote unspecified constants, possibly different at each occurrence. They will in general depend on the Pólya urn, i.e., on q, X_0 and the distribution of ξ_i, $i \in [q]$, but the do not depend on n. We similarly use C_p for constants that also may depend on the power $p \geqslant 2$.

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

$$A := (a_j \mathbb{E}\,\xi_{ji})_{i,j=1}^q. \tag{2.4}$$

Note that, for convenience and following [11] and [13], 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. (We may unfortunately have contributed to notational confusion by this choice in [11].) It is well-known that the intensity matrix A and in particular its eigenvalues and eigenvectors have a central role for asymptotical results.

2.4. **Eigenvalues and spectral decomposition.** 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_λ, such that $A - \lambda I$ is a nilpotent operator on E_λ; here I is the identity matrix and λ ranges over the set $\sigma(A)$ of eigenvalues of A. (In the sequel, λ will always denote an eigenvalue.) In other words, there exist projections P_λ, $\lambda \in \sigma(A)$, that commute with A and satisfy

$$\sum_{\lambda\in\sigma(A)} P_\lambda = I, \tag{2.5}$$

$$AP_\lambda = P_\lambda A = \lambda P_\lambda + N_\lambda, \tag{2.6}$$

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 $\nu_\lambda \geqslant 0$ be the integer such that $N_\lambda^{\nu_\lambda} \neq 0$ but $N_\lambda^{\nu_\lambda+1} = 0$. (Equivalently, in the Jordan normal form of A, the largest Jordan block with λ on the diagonal has size $\nu_\lambda + 1$.) Hence $\nu_\lambda = 0$ if and only if $N_\lambda = 0$, and this happens for all λ if and only if A is diagonalizable.

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: $\operatorname{Re}\lambda_1 \geqslant \operatorname{Re}\lambda_2 \geqslant \dots$, and furthermore, when the real parts are equal, in order of decreasing $\nu_j := \nu_{\lambda_j}$. In particular, if $\lambda_1 > \operatorname{Re}\lambda_2$, then $\nu_j \leqslant \nu_2$ for every eigenvalue λ_j with $\operatorname{Re}\lambda_j = \operatorname{Re}\lambda_2$.

Recall that the urn is called *small* if $\operatorname{Re}\lambda_2 \leqslant \frac{1}{2}\lambda_1$ and *large* if $\operatorname{Re}\lambda_2 > \frac{1}{2}\lambda_1$.

In the balanced case, by (2.4) and (2.3),

$$a'A = \Bigl(\sum_{i=1}^q a_i (A)_{ij}\Bigr)_j = \Bigl(\sum_{i=1}^q a_i a_j \mathbb{E}\,\xi_{ji}\Bigr)_j = \bigl(a_j \mathbb{E}(a\cdot\xi_j)\bigr)_j = ba', \tag{2.7}$$

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.,

$$\lambda_1 = b. \tag{2.8}$$

This is a very weak assumption; see [13, Appendix A] where it is shown that there are counterexamples but only in a trivial manner. In particular, the following is shown there.

Lemma 2.2 ([13, Lemma A.1])**.** *If the Pólya urn is tenable and balanced, and moreover any colour has a nonzero probability of ever appearing in the urn, then* $\operatorname{Re}\lambda \leqslant b$ *for every* $\lambda \in \sigma(A)$*, and, furthermore, if* $\operatorname{Re}\lambda = b$ *then* $\nu_\lambda = 0$*. We may thus assume* $\lambda_1 = b$*.*

The eigenvalue b may be multiple; a well-known example is the classical Pólya urn for which $A = bI$, and thus all q eigenvalues are equal to b. In most other applications, the eigenvalue b is simple. This implies that the corresponding left and right eigenspaces are 1-dimensional, and thus the corresponding left and right eigenvectors u_1 and v_1 are unique up to constant factors. By (2.7), a' is a left eigenvector so we may then take $u_1 = a$. Furthermore, we have the following general result from linear algebra; for completeness we give a proof in Appendix B since we do not know a good reference.

Lemma 2.3. *Suppose that the eigenvalue* $\lambda_1 = b$ *is simple. Then there is a unique right eigenvector* v_1 *with*

$$u_1 \cdot v_1 = a \cdot v_1 = 1. \tag{2.9}$$

Furthermore, the projection P_{λ_1} *is given by*

$$P_{\lambda_1} = v_1 u_1'. \tag{2.10}$$

Consequently, for any vector $v \in \mathbb{C}^q$*,*

$$P_{\lambda_1} v = v_1 u_1' v = v_1 a' v = (a \cdot v) v_1. \tag{2.11}$$

3. Main results

We state here our main results, using the notation and assumptions above. Proofs are given in Section 4. We begin with a general upper bound for the moments.

Theorem 3.1. *Assume that the Pólya urn is tenable and balanced with* $\lambda_1 = b$*. Let* $p \geqslant 2$ *and suppose that*

$$\mathbb{E}\,|\xi_{ij}|^p < \infty \qquad \text{for all } i,j \in [q]. \tag{3.1}$$

Then, for every $n \geqslant 2$*,*

$$\big\|X_n - \mathbb{E}\,X_n\big\|_p \leqslant \begin{cases} C_p n^{1/2}, & \operatorname{Re}\lambda_2 < \lambda_1/2, \\ C_p n^{1/2}(\log n)^{\nu_2+\frac12}, & \operatorname{Re}\lambda_2 = \lambda_1/2, \\ C_p n^{\operatorname{Re}\lambda_2/\lambda_1}(\log n)^{\nu_2}, & \operatorname{Re}\lambda_2 > \lambda_1/2, \end{cases} \tag{3.2}$$

for some constant C_p *not depending on* n*.*

As said in the introduction, in many cases the asymptotic distribution of the urn is known. Furthermore, for an irreducible urn, under some technical conditions, the bound in (3.2) equals (apart from the constant C_p) the right normalizing factor, see for example [11, Theorems 3.22–24]. In particular, for a small irreducible urn (again under some conditions), we have asymptotic normality as in (3.3) or (3.4) below. The following theorem shows that then also all moments (ordinary and absolute) converge.

Theorem 3.2. *Assume that the Pólya urn is tenable and balanced, and suppose that* (3.1) *holds for every* $p \geqslant 1$*.*

(i) *Assume that* $\operatorname{Re}\lambda_2 < \lambda_1/2$ *and that, as* $n \to \infty$*, we have asymptotic normality*

$$\frac{X_n - \mathbb{E}\,X_n}{\sqrt{n}} \xrightarrow{\mathrm{d}} N(0,\Sigma) \tag{3.3}$$

for some matrix Σ*. Then* (3.3) *holds with convergence of all moments. In particular, the covariance matrix* $\operatorname{Cov}[X_n]/n \to \Sigma$*.*

(ii) *Assume that* $\operatorname{Re}\lambda_2 = \lambda_1/2$ *and that, as* $n \to \infty$*, we have asymptotic normality*

$$\frac{X_n - \mathbb{E}\,X_n}{\sqrt{n(\log n)^{2\nu_2+1}}} \xrightarrow{\mathrm{d}} N(0,\Sigma) \tag{3.4}$$

for some matrix Σ*. Then* (3.4) *holds with convergence of all moments. In particular, the covariance matrix* $\operatorname{Cov}[X_n]/\big(n(\log n)^{2\nu_2+1}\big) \to \Sigma$*.*

We may also estimate different spectral projections separately and get a sharper result than Theorem 3.1.

Theorem 3.3. *Assume that the Pólya urn is tenable and balanced. Let* $p \geqslant 2$ *and suppose that* (3.1) *holds. Let* λ *be an eigenvalue of* A*. Then, for every* $n \geqslant 2$*,*

$$\big\|P_\lambda(X_n - \mathbb{E}\,X_n)\big\|_p \leqslant \begin{cases} C_p n^{1/2}, & \operatorname{Re}\lambda < b/2,\\ C_p n^{1/2}(\log n)^{\nu_\lambda+\frac12}, & \operatorname{Re}\lambda = b/2,\\ C_p n^{\operatorname{Re}\lambda/b}(\log n)^{\nu_\lambda}, & \operatorname{Re}\lambda > b/2, \end{cases} \tag{3.5}$$

for some constant C_p *not depending on* n*.*

As in Theorem 3.2, one can often combine Theorem 3.3 with a central limit result for a specific components $P_\lambda(X_n - \mathbb{E}\,X_n)$ (or for suitable linear combinations $u\cdot(X_n - \mathbb{E}\,X_n)$) and obtain moment convergence in the latter results, cf. [11, Remark 3.25] and [1, Theorem 3]; we leave the details to the reader.

For the special case $\lambda = \lambda_1$, and further assuming that this eigenvalue is simple, we have as a complement the following almost trivial result.

Theorem 3.4. *Assume that the Pólya urn is tenable and balanced and that* $\lambda_1 = b$ *is a simple eigenvalue. Then*

$$P_{\lambda_1}(X_n - \mathbb{E}\,X_n) = 0. \tag{3.6}$$

When $\operatorname{Re}\lambda \leqslant b/2$, we have under quite general condition asymptotic normality of $P_\lambda(X_n - \mathbb{E}\,X_n)$, for example as a consequence of [11, Theorem 3.15]. In such cases, we obtain from Theorem 3.3 also moment convergence of $P_\lambda(X_n - \mathbb{E}\,X_n)$, by the same argument as in the proof of Theorem 3.2.

Remark 3.5. The assumption (3.1) on finite moments of the replacements is a very weak restriction in the balanced case studied here. For example, for a balanced and tenable urn such that every colour may appear with positive probability (which we may assume without loss of generality) and all activities $a_i > 0$, we necessarily have all ξ_{ij} bounded by some constant, and thus (3.1) holds, see [13, Remark 2.5]. Nevertheless, we note for completeness that if we assume only that (3.1) holds for a single $p \geqslant 2$, then our proof of Theorem 3.2 in Section 4 yields only moment convergence for moments of order strictly less that p; however, we show in Appendix A a modification of the argument which yields convergence also of moments of order p in this case.

4. Proofs of main results

We prove here the main results in Section 3, using some lemmas postponed to the following sections. We begin with standard arguments, partly copying from [13].

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

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

and

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

the total weight (activity) of the urn. We note first that the assumption that the urn is balanced implies $a \cdot \Delta X_n = b$ a.s., and thus (4.1)–(4.2) imply that w_n is deterministic with

$$w_n = w_0 + nb, \tag{4.3}$$

where the initial weight $w_0 = a \cdot X_0$.

Next, let $\mathcal{F}_n$ be the σ-field generated by $X_1, \dots, X_n$. Then, by the definition of the urn,

$$\mathbb{P}\bigl(I_{n+1} = j \mid \mathcal{F}_n\bigr) = \frac{a_j X_{nj}}{w_n} \tag{4.4}$$

and consequently, recalling (2.4),

$$\begin{aligned} \mathbb{E}\bigl(\Delta X_n \mid \mathcal{F}_n\bigr) &= \sum_{j=1}^{q} \mathbb{P}\bigl(I_{n+1} = j \mid \mathcal{F}_n\bigr) \mathbb{E}\, \xi_j = \frac{1}{w_n} \sum_{j=1}^{q} a_j X_{nj} \, \mathbb{E}\, \xi_j \\ &= \frac{1}{w_n} \biggl(\sum_{j=1}^{q} (A)_{ij} X_{nj} \biggr)_i = \frac{1}{w_n} A X_n. \end{aligned} \tag{4.5}$$

Define

$$Y_n := \Delta X_{n-1} - \mathbb{E}\bigl(\Delta X_{n-1} \mid \mathcal{F}_{n-1}\bigr), \qquad n \geqslant 1. \tag{4.6}$$

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

$$\mathbb{E}\bigl(Y_n \mid \mathcal{F}_{n-1}\bigr) = 0. \tag{4.7}$$

In other words, $(Y_n)_1^\infty$ is a martingale difference sequence.

Furthermore, similarly to (4.5) and using the assumption (3.1),

$$\mathbb{E}\bigl(|\Delta X_n|^p \mid \mathcal{F}_n\bigr) = \sum_{j=1}^{q} \mathbb{P}\bigl(I_{n+1} = j \mid \mathcal{F}_n\bigr) \mathbb{E}\, |\xi_j|^p \leqslant C_p \quad \text{a.s.} \tag{4.8}$$

Hence, $\mathbb{E}\, |\Delta X_n|^p \leqslant C_p$, or equivalently

$$\|\Delta X_n\|_p \leqslant C_p, \tag{4.9}$$

which by (4.6) implies

$$\|Y_n\|_p \leqslant C_p. \tag{4.10}$$

By (4.1), (4.6) and (4.5),

$$X_{n+1} = X_n + Y_{n+1} + w_n^{-1} A X_n = \bigl(I + w_n^{-1} A\bigr) X_n + Y_{n+1}. \tag{4.11}$$

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

$$X_n = \prod_{k=0}^{n-1} \bigl(I + w_k^{-1}A\bigr)X_0 + \sum_{\ell=1}^{n} \prod_{k=\ell}^{n-1} \bigl(I + w_k^{-1}A\bigr)Y_\ell, \tag{4.12}$$

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

We define the (deterministic) matrix products

$$F_{i,j} := \prod_{i\leqslant k<j} \bigl(I + w_k^{-1}A\bigr), \qquad 0 \leqslant i \leqslant j, \tag{4.13}$$

and write (4.12) as

$$X_n = F_{0,n}X_0 + \sum_{\ell=1}^{n} F_{\ell,n}Y_\ell. \tag{4.14}$$

Taking the expectation we find, since $\mathbb{E}\,Y_\ell = 0$ by (4.7), and the $F_{i,j}$ and X_0 are nonrandom,

$$\mathbb{E}\,X_n = F_{0,n}X_0. \tag{4.15}$$

Hence, (4.14) can also be written

$$X_n - \mathbb{E}\,X_n = \sum_{\ell=1}^{n} F_{\ell,n}Y_\ell. \tag{4.16}$$

Proof of Theorem 3.3. It follows from (4.16) that

$$P_\lambda(X_n - \mathbb{E}\,X_n) = \sum_{\ell=1}^{n} P_\lambda F_{\ell,n}Y_\ell. \tag{4.17}$$

We now use Lemma 5.1 below and conclude using (4.10) that

$$\bigl\|P_\lambda(X_n - \mathbb{E}\,X_n)\bigr\|_p \leqslant C_p\Bigl(\sum_{\ell=1}^{n}\bigl\|P_\lambda F_{\ell,n}\bigr\|^2\Bigr)^{1/2}. \tag{4.18}$$

The result follows by Lemma 6.2. □

Proof of Theorem 3.4. By (4.2) and (4.3), $a \cdot X_n = w_n$ is deterministic, and thus

$$a \cdot (X_n - \mathbb{E}\,X_n) = a \cdot X_n - E(a \cdot X_n) = 0. \tag{4.19}$$

The result now follows from (2.11) in Lemma 2.3. □

Proof of Theorem 3.1. First, if $\operatorname{Re}\lambda_2 = \lambda_1 = b$, then we simply use Minkowski's inequality, which by (4.9) yields

$$\|X_n\|_p \leqslant \|X_0\|_p + \sum_{i=0}^{n-1} \|\Delta X_i\|_p \leqslant C_p + C_p n \leqslant C_p n. \tag{4.20}$$

Hence, $\|\,\mathbb{E}\,X_n\|_p \leqslant \|X_n\|_p \leqslant C_p n$ and $\|X_n - \mathbb{E}\,X_n\| \leqslant C_p n$, which is (3.2) in the case $\operatorname{Re}\lambda_2 = \lambda_1$.

In the rest of the proof, suppose instead that $\operatorname{Re}\lambda_2 < \lambda_1$. In particular (since eigenvalues are counted with multiplicities), λ_1 is a simple eigenvalue.

Thus Theorem 3.4 applies and shows $P_{\lambda_1}(X_n - \mathbb{E}\,X_n) = 0$. The decomposition (2.5) thus yields

$$X_n - \mathbb{E}\,X_n = \sum_{\lambda \neq \lambda_1} P_\lambda(X_n - \mathbb{E}\,X_n), \tag{4.21}$$

and Minkowski's inequality yields

$$\|X_n - \mathbb{E}\,X_n\|_p \leqslant \sum_{\lambda \neq \lambda_1} \|P_\lambda(X_n - \mathbb{E}\,X_n)\|_p. \tag{4.22}$$

We estimate the terms on the right-hand side by Theorem 3.3; the contribution from $\lambda = \lambda_2$ dominates all others, and we obtain (3.2). □

Proof of Theorem 3.2. It follows from Theorem 3.3 that for every $p \geqslant 1$, the L^p norms of the left-hand sides of (3.3) and (3.4) are bounded as $n \to \infty$. As is well-known this implies that for every $p \geqslant 1$, the pth powers $|X_n - \mathbb{E}\,X_n|^p$ are uniformly integrable, and hence convergence of moments follows from the assumed convergence in distribution. □

5. A martingale inequality

We used above the following martingale inequality, which is a simple consequence of Burkholder's inequality for the square function. Since we do not know a reference where this inequality is stated in the form below, we give a complete proof. Recall that a martingale difference sequence is a sequence $(Y_i)_1^\infty$ of random variables such that the sequence $\sum_{i=1}^n Y_i$, $n \geqslant 1$, is a martingale with respect to some sequence of σ-fields $\mathcal{F}_n$. (The σ-fields $\mathcal{F}_n$ will be fixed below.)

Lemma 5.1. *Let $p \geqslant 2$ and let Y_i, $i \geqslant 1$, be a martingale difference sequence of random vectors in $\mathbb{C}^q$ such that $\sup_i \|Y_i\|_p < \infty$. Then, for any sequence of (nonrandom) $q \times q$ matrices $(A_i)_{i=1}^\infty$ such that $\sum_{i=1}^\infty \|A_i\|^2 < \infty$, the sum $\sum_{i=1}^\infty A_i Y_i$ converges a.s. and in L^p, and*

$$\Bigl\|\sum_{i=1}^\infty A_i Y_i\Bigr\|_p \leqslant C_p \left(\sum_{i=1}^\infty \|A_i\|^2\right)^{1/2} \sup_i \|Y_i\|_p. \tag{5.1}$$

Here C_p is a constant that depends on p and q only.

Proof. $X_n := \sum_{i=1}^n A_i Y_i$, $n \geqslant 0$, is a martingale, and its square function is

$$S_n(X) := \left(\sum_{i=1}^n |X_i - X_{i-1}|^2\right)^{1/2} = \left(\sum_{i=1}^n |A_i Y_i|^2\right)^{1/2}. \tag{5.2}$$

Let $B := \sup_i \|Y_i\|_p$. Minkowski's inequality yields, for any $n \geqslant 1$,

$$\begin{aligned}
\|S_n(X)\|_p^2 = \bigl\|S_n(X)^2\bigr\|_{p/2} &= \Bigl\|\sum_{i=1}^n |A_i Y_i|^2\Bigr\|_{p/2} \\
&\leqslant \sum_{i=1}^n \bigl\||A_i Y_i|^2\bigr\|_{p/2} \leqslant \sum_{i=1}^n \|A_i\|^2 \||Y_i|^2\|_{p/2} = \sum_{i=1}^n \|A_i\|^2 \|Y_i\|_p^2 \\
&\leqslant B^2 \sum_{i=1}^n \|A_i\|^2 \leqslant B^2 \sum_{i=1}^\infty \|A_i\|^2,
\end{aligned} \tag{5.3}$$

i.e., $\|S_n(X)\|_p \leqslant B\bigl(\sum_{i=1}^{\infty}\|A_i\|^2\bigr)^{1/2}$. We combine this with Burkholder's inequality

$$\|X_n\|_p \leqslant C_p\|S(X)_n\|_p \tag{5.4}$$

(valid for any martingale and any $p \geqslant 1$, with some constant C_p depending on p only), see [7, Theorem 9] or e.g. [10, Theorem 10.9.5]. This yields

$$\|X_n\|_p \leqslant C_p\|S_n(X)\|_p \leqslant C_pB\left(\sum_{i=1}^{\infty}\|A_i\|^2\right)^{1/2}. \tag{5.5}$$

Consequently, the martingale $(X_n)_1^{\infty}$ is L^p bounded, and thus converges a.s. and in L^p to a limit, which satisfies (5.1) by (5.5). ☐

Remark 5.2. We have in the proof above used a vector-valued version of Burkholder's inequality (5.4); this follows immediately (with a constant C_p depending on q) from the scalar-valued version in the references above applied to each component. In fact (for $p \geqslant 2$), (5.4) holds with $C_p = p-1$ independent of the dimension q, and, in fact, more generally for Hilbert-space-valued martingales, see [8, Theorem 3.3].

6. Matrix estimates

We prove here some matrix estimates used in the proofs above. In this section, we may as well be general and let A be any complex $q \times q$ matrix. Let P_λ, $\lambda \in \sigma(A)$, be its spectral projections as in (2.5)–(2.6), and let $\nu_\lambda + 1$ be the dimension of the largest Jordan block for the eigenvalue λ.

Furthermore, we assume that w_0 and $b > 0$ are some given positive numbers. We then define $w_n := w_0 + nb > 0$ as in (4.3), and define the matrices $F_{i,j}$ by (4.13).

Lemma 6.1. *With notations as above, for every eigenvalue $\lambda \in \sigma(A)$,*

$$\|P_\lambda F_{i,j}\| \leqslant C\left(\frac{j}{i}\right)^{\operatorname{Re}\lambda/b}\left(1+\log\frac{j}{i}\right)^{\nu_\lambda}, \qquad 1 \leqslant i \leqslant j < \infty. \tag{6.1}$$

for some constant C not depending on i and j.

Proof. We change basis in $\mathbb{C}^q$ so that A is reduced to Jordan normal form. (This may change the matrix norms, but at most by some multiplicative constants, which are incorporated in the final C and do not affect the result.) In this basis, A has one or several Jordan blocks with diagonal element λ. Multiplying by P_λ kills all Jordan blocks with eigenvalue $\neq \lambda$, and as a result it suffices to consider a single Jordan block with eigenvalue λ. We may thus assume that $A = \lambda I + N$ where N is nilpotent, cf. (2.6); more precisely we have $N^{\nu_\lambda+1} = 0$ by the definition of ν_λ. Assume also that i is so large that $w_i \geqslant 2|\lambda|$, say, and thus in particular $w_k + \lambda \neq 0$ for $k \geqslant i$. In this case, by (4.13),

$$\begin{aligned} P_\lambda F_{i,j} &= \prod_{i\leqslant k<j}\bigl(I + w_k^{-1}(\lambda I + N)\bigr) && (6.2)\\ &= \prod_{i\leqslant k<j}(1+\lambda/w_k)\prod_{i\leqslant k<j}\bigl(I+(w_k+\lambda)^{-1}N\bigr). && (6.3)\end{aligned}$$

The first product on the right-hand side of (6.3) is a complex number which can be estimated by (since we assume $w_k \geqslant w_i \geqslant 2|\lambda|$)

$$\begin{aligned}\prod_{i\leqslant k<j}(1+\lambda/w_k) &= \exp\Bigl(\sum_{i\leqslant k<j}\log(1+\lambda/w_k)\Bigr)\\ &= \exp\Bigl(\sum_{i\leqslant k<j}\Bigl(\frac{\lambda}{w_k}+O\Bigl(\frac{\lambda^2}{w_k^2}\Bigr)\Bigr)\Bigr)\\ &= \exp\Bigl(\sum_{i\leqslant k<j}\Bigl(\frac{\lambda}{kb}+O\Bigl(\frac{1}{k^2}\Bigr)\Bigr)\Bigr)\\ &= \exp\Bigl(\frac{\lambda}{b}\sum_{i\leqslant k<j}\frac{1}{k}+O(1)\Bigr). \end{aligned} \tag{6.4}$$

Hence,

$$\begin{aligned}\Bigl|\prod_{i\leqslant k<j}(1+\lambda/w_k)\Bigr| &= \exp\Bigl(\frac{\operatorname{Re}\lambda}{b}\bigl(\log j-\log i\bigr)+O(1)\Bigr)\\ &\leqslant C\left(\frac{j}{i}\right)^{\operatorname{Re}\lambda/b}. \end{aligned} \tag{6.5}$$

We turn to the final (matrix) product in (6.3) and expand it into a polynomial $\sum_\ell a_\ell N^\ell$ in N. The coefficient a_ℓ of N^ℓ is a sum of the product $\prod_{j=1}^{\ell}(w_{k_j}+\lambda)^{-1}$ over ℓ-tuples $k_1<\dots<k_\ell$ of indices. Hence,

$$|a_\ell| \leqslant \Bigl(\sum_{i\leqslant k<j}|w_k+\lambda|^{-1}\Bigr)^{\ell}. \tag{6.6}$$

We have, similarly to (6.4),

$$\sum_{i\leqslant k<j}|w_k+\lambda|^{-1} = \sum_{i\leqslant k<j}\Bigl(\frac{1}{bk}+O\Bigl(\frac{1}{k^2}\Bigr)\Bigr) \leqslant C\Bigl(1+\log\frac{j}{i}\Bigr). \tag{6.7}$$

Since $N^\ell=0$ for $\ell>\nu_\lambda$, we only have to consider $a_\ell N^\ell$ for $\ell\leqslant\nu_\lambda$. Hence, (6.6)–(6.7) imply

$$\Bigl\|\prod_{i\leqslant k<j}\bigl(I+(w_k+\lambda)^{-1}N\bigr)\Bigr\| \leqslant \sum_{\ell=0}^{\nu_\lambda}C\Bigl(1+\log\frac{j}{i}\Bigr)^{\ell} \leqslant C\Bigl(1+\log\frac{j}{i}\Bigr)^{\nu_\lambda}. \tag{6.8}$$

Combining (6.3), (6.5) and (6.8) we obtain (6.1).

This proof of (6.1) assumed that $i\geqslant i_0$ for some i_0 such that $w_{i_0}\geqslant 2|\lambda|$. However, for $i<i_0$ we may use (6.1) with $i=i_0$ and multiply by the missing factors in (6.2) (which are bounded) to conclude that (6.1) holds for all $1\leqslant i\leqslant j<\infty$. □

Lemma 6.2. *With notations as above, for every eigenvalue $\lambda\in\sigma(A)$ and $n\geqslant 2$,*

$$\sum_{i=1}^{n}\|P_\lambda F_{i,n}\|^2 \leqslant \begin{cases} Cn, & \operatorname{Re}\lambda<b/2,\\ Cn\log^{1+2\nu_\lambda} n, & \operatorname{Re}\lambda=b/2,\\ Cn^{2\operatorname{Re}\lambda/b}\log^{2\nu_\lambda} n, & \operatorname{Re}\lambda>b/2, \end{cases} \tag{6.9}$$

for some constant C not depending on n.

Proof. This follows from Lemma 6.1 by simple calculations. Let $\gamma := \operatorname{Re}\lambda/b$. If $\gamma < \frac{1}{2}$, we obtain from Lemma 6.1, by comparing the sum to a convergent integral,

$$\sum_{i=1}^{n} \|P_\lambda F_{i,n}\|^2 \leqslant C \sum_{i=1}^{n} \Bigl(\frac{i}{n}\Bigr)^{-2\gamma} \Bigl(1+\log\frac{n}{i}\Bigr)^{2\nu_\lambda}$$

$$\leqslant Cn \int_0^1 x^{-2\gamma}\Bigl(1+\log\frac{1}{x}\Bigr)^{2\nu_\lambda}\,\mathrm{d}x = Cn. \tag{6.10}$$

If $\gamma = \frac{1}{2}$ we obtain instead

$$\sum_{i=1}^{n} \|P_\lambda F_{i,n}\|^2 \leqslant Cn \sum_{i=1}^{n} i^{-1}\bigl(1+\log n\bigr)^{2\nu_\lambda} \leqslant Cn(\log n)^{1+2\nu_\lambda}. \tag{6.11}$$

Finally, if $\gamma > \frac{1}{2}$, we obtain

$$\sum_{i=1}^{n} \|P_\lambda F_{i,n}\|^2 \leqslant C \sum_{i=1}^{n} \Bigl(\frac{n}{i}\Bigr)^{2\gamma} \bigl(\log n\bigr)^{2\nu_\lambda} \leqslant Cn^{2\gamma}\bigl(\log n\bigr)^{2\nu_\lambda} \sum_{i=1}^{\infty} i^{-2\gamma}, \tag{6.12}$$

where the final sum is convergent. ∎

Appendix A. Uniform integrability

The purpose of this appendix is to prove the following version of Theorem 3.2, useful in the (unusual) case of a balanced urn where the increments ξ_i have some but not all moments finite.

Theorem A.1. *Assume that the conditions of Theorem 3.2*(i) *or* (ii) *holds, except that* (3.1) *is assumed only for a single $p \geqslant 2$. Then* (3.3) *or* (3.4) *holds with convergence of all moments of order less than or equal to p.*

Proof. We argue as in the proof of Theorem 3.3 above, but use Lemma A.2 below instead of Lemma 5.1. It follows that if L_n is the random vector on the left-hand side of (3.5) and r_n is the function of n on the right-hand side, then $|L_n/r_n|^p$ forms a uniformly integrable sequence. It follows as in the proof of Theorem 3.1 that the same holds for (3.2); in other words, the left-hand side of (3.3) or (3.4) has uniformly integrable pth powers. Hence convergence of pth moments follows from the assumed convergence in distribution. ∎

The proof above uses the following lemma, which was shown to me by an anonymous referee of [13]. We do not know any reference, so we give a complete proof.

Lemma A.2. *Let $p \geqslant 2$ and let Y_i, $i \geqslant 1$, be a martingale difference sequence of random vectors in $\mathbb{C}^q$ such that the variables $|Y_i|^p$, $i \geqslant 1$, are uniformly integrable. Then the collection of random variables*

$$\Bigl\{\Bigl|\sum_{i=1}^{\infty} A_i Y_i\Bigr|^p : A_i \in \mathbb{C}^{q\times q},\ \sum_{i=1}^{\infty} \|A_i\|^2 \leqslant 1\Bigr\} \tag{A.1}$$

is uniformly integrable. (The first sum in (A.1) *converges a.s.)*

Note that the a.s. convergence of $\sum_i A_i Y_i$ follows from Lemma 5.1, which also shows that the L^p-norms of these sums are bounded. We obtain Lemma A.2 from Lemma 5.1 by a simple truncation argument.

Proof of Lemma A.2. Let $\varepsilon > 0$. By assumption, there exists $M = M(\varepsilon)$ such that

$$\mathbb{E}\big|Y_i \mathbf{1}_{|Y_i|>M}\big|^p < \varepsilon^p, \qquad i \geqslant 1. \tag{A.2}$$

Let

$$Y_i' := Y_i \mathbf{1}_{|Y_i|\leqslant M} - \mathbb{E}\big(Y_i \mathbf{1}_{|Y_i|\leqslant M} \mid \mathcal{F}_{i-1}\big), \tag{A.3}$$

$$Y_i'' := Y_i \mathbf{1}_{|Y_i|> M} - \mathbb{E}\big(Y_i \mathbf{1}_{|Y_i|> M} \mid \mathcal{F}_{i-1}\big). \tag{A.4}$$

Then $Y_i = Y_i' + Y_i''$. Furthermore,

$$|Y_i'| \leqslant 2M \qquad \text{a.s.}, \tag{A.5}$$

$$\|Y_i''\|_p \leqslant 2\big\|Y_i \mathbf{1}_{|Y_i|>M}\big\|_p < 2\varepsilon. \tag{A.6}$$

Let $r := p+1$ (any fixed $r > p$ will do), and note that (A.5) implies

$$\|Y_i'\|_r \leqslant 2M. \tag{A.7}$$

Let $(A_i)_1^\infty$ be any sequence of matrices with $\sum_i \|A_i\|^2 \leqslant 1$. Then Lemma 5.1 yields, together with (A.7) and (A.6),

$$\Big\|\sum_i A_i Y_i'\Big\|_r \leqslant 2C_r M, \tag{A.8}$$

$$\Big\|\sum_i A_i Y_i''\Big\|_p \leqslant 2C_p \varepsilon, \tag{A.9}$$

Let $\delta > 0$ and let $\mathcal{E}$ be any event with $\mathbb{P}(\mathcal{E}) \leqslant \delta$. We have $\sum_i A_i Y_i = \sum_i A_i Y_i' + \sum_i A_i Y_i''$, and thus (crudely),

$$\Big|\sum_i A_i Y_i\Big|^p \leqslant 2^p \Big|\sum_i A_i Y_i'\Big|^p + 2^p \Big|\sum_i A_i Y_i''\Big|^p. \tag{A.10}$$

Hence, using Hölder's inequality with $s := (r/p)' = r/(r-p)$, and (A.8)–(A.9),

$$\begin{aligned}
\mathbb{E}\left(\Big|\sum_i A_i Y_i\Big|^p \mathbf{1}_{\mathcal{E}}\right) &\leqslant 2^p\, \mathbb{E}\left(\Big|\sum_i A_i Y_i'\Big|^p \mathbf{1}_{\mathcal{E}}\right) + 2^p\, \mathbb{E}\left(\Big|\sum_i A_i Y_i''\Big|^p \mathbf{1}_{\mathcal{E}}\right) \\
&\leqslant 2^p \left\|\Big|\sum_i A_i Y_i'\Big|^p\right\|_{r/p} \|\mathbf{1}_{\mathcal{E}}\|_s + 2^p\, \mathbb{E}\left(\Big|\sum_i A_i Y_i''\Big|^p\right) \\
&= 2^p \Big\|\sum_i A_i Y_i'\Big\|_r^p \mathbb{P}(\mathcal{E})^{1/s} + 2^p \Big\|\sum_i A_i Y_i''\Big\|_p^p \\
&\leqslant 2^p (2C_r M)^p \delta^{1/s} + 2^p (2C_p\varepsilon)^p.
\end{aligned} \tag{A.11}$$

For any $\eta > 0$, we can make the right-hand side of (A.11) $< \eta + \eta$ by first choosing ε and then δ small enough. Consequently,

$$\lim_{\delta\to 0} \sup_{\mathbb{P}(\mathcal{E})\leqslant\delta,\, \sum_i \|A_i\|^2\leqslant 1} \mathbb{E}\left(\Big|\sum_i A_i Y_i\Big|^p \mathbf{1}_{\mathcal{E}}\right) = 0. \tag{A.12}$$

Finally, the assumption implies $\sup_i \mathbb{E}\,|Y_i|^p < \infty$, and thus another application of Lemma 5.1 yields $\sup_{\sum_i \|A_i\|^2\leqslant 1} \mathbb{E}\big|\sum_i A_i Y_i\big|^p < \infty$, which together with (A.12) shows the uniform integrability of (A.1). $\square$

Appendix B. Proof of Lemma 2.3

Proof. Consider an eigenvalue $\lambda \neq \lambda_1 = b$. Then (2.6) shows that $(A - \lambda I)P_\lambda = N_\lambda$, and thus, since $\nu_\lambda < q$, $(A-\lambda I)^q P_\lambda = ((A-\lambda I)P_\lambda)^q = 0$. (Recall that P_λ is a projection and that it commutes with A.) Hence, for any vector $v \in \mathbb{C}^q$, since $u_1(A-\lambda I) = (\lambda_1 - \lambda)u_1$,

$$0 = u_1'(A-\lambda I)^q P_\lambda v = (\lambda_1-\lambda)^q u_1 P_\lambda v, \tag{B.1}$$

and consequently

$$u_1' P_\lambda v = 0. \tag{B.2}$$

This says that u_1 is orthogonal to the generalized eigenspace $E_\lambda = P_\lambda \mathbb{C}^q$ for every $\lambda \neq \lambda_1$. Since $\mathbb{C}^q = \bigoplus_\lambda E_\lambda$ and $u_1 \neq 0$, it follows that u_1 is not orthogonal to $E_{\lambda_1} = \{zv_1 : z \in \mathbb{C}\}$. Hence $u_1 \cdot v_1 \neq 0$ and we may choose v_1 such that (2.9) holds.

Since P_{λ_1} is a projection onto the eigenspace E_{λ_1} spanned by v_1 we have

$$P_{\lambda_1} v = (u \cdot v) v_1, \qquad v \in \mathbb{C}^q \tag{B.3}$$

for some vector u with $u \cdot v_1 = 1$. Furthermore,

$$u_1' P_{\lambda_1} A = u_1' A P_{\lambda_1} = \lambda_1 u_1' P_{\lambda_1} \tag{B.4}$$

and thus $u_1' P_{\lambda_1}$ is a left eigenvector of A with eigenvalue λ_1 and thus a multiple of u_1. Since P_{λ_1} is a projection, it follows that $u_1' P_{\lambda_1} = u_1'$. Hence (B.3) implies that, for any $v \in \mathbb{C}^q$,

$$u_1' v = u_1' P_{\lambda_1} v = (u \cdot v)(u_1 \cdot v_1) = u \cdot v. \tag{B.5}$$

Consequently, $u = u_1$; thus (B.3) yields (2.11), and thus also (2.10). ☐

References

[1] Krishna B. Athreya and Samuel Karlin: Embedding of urn schemes into continuous time Markov branching processes and related limit theorems. *Ann. Math. Statist.* **39** (1968), 1801–1817.

[2] A. Bagchi and A. K. Pal: Asymptotic normality in the generalized Pólya–Eggenberger urn model, with an application to computer data structures. *SIAM J. Algebraic Discrete Methods* **6** (1985), no. 3, 394–405.

[3] Zhi-Dong Bai and Feifang Hu: Asymptotic theorems for urn models with nonhomogeneous generating matrices. *Stochastic Process. Appl.* **80** (1999), no. 1, 87–101.

[4] Zhi-Dong Bai and Feifang Hu: Asymptotics in randomized urn models. *Ann. Appl. Probab.* **15** (2005), no. 1B, 914–940.

[5] Sergei N. Bernstein: Nouvelles applications des grandeurs aléatoires presqu'indépendantes. (Russian) *Izv. Akad. Nauk SSSR Ser. Mat.* **4** (1940), no. 2, 137–150.

[6] Sergei N. Bernstein: Sur un problème du schéma des urnes à composition variable. *C. R. (Doklady) Acad. Sci. URSS (N.S.)* **28** (1940), 5–7.

[7] Donald L. Burkholder: Martingale transforms. *Ann. Math. Statist.* **37** (1966), 1494–1504.

[8] Donald L. Burkholder: Explorations in martingale theory and its applications. *École d'Été de Probabilités de Saint-Flour XIX—1989*, 1–66. Lecture Notes in Math. **1464**, Springer-Verlag, Berlin, 1991

[9] F. Eggenberger and G. Pólya: Über die Statistik verketteter Vorgänge. *Zeitschrift Angew. Math. Mech.* **3** (1923), no. 4, 279–289.

[10] Allan Gut: *Probability: A Graduate Course*, 2nd ed., Springer, New York, 2013.

[11] Svante Janson: Functional limit theorems for multitype branching processes and generalized Pólya urns. *Stoch. Process. Appl.* **110** (2004), 177–245.

[12] Svante Janson: Limit theorems for triangular urn schemes. *Probab. Theory Rel. Fields* **134** (2005), 417–452.

[13] Svante Janson: Mean and variance of balanced Pólya urns. *Adv. in Appl. Probab.* **52**:4 (2020), 1224–1248.

[14] Svante Janson: Almost sure and moment convergence for triangular Pólya urns. Preprint, 2024. `arXiv:2402.01299`

[15] Svante Janson and Nicholas Pouyanne: Moment convergence of balanced Pólya processes. *Electron. J. Probab.* **23** (2018), Paper No. 34, 13 pp.

[16] Hosam M. Mahmoud: *Pólya Urn Models.* CRC Press, Boca Raton, FL, 2009.

[17] Andrey A. Markov: Sur quelques formules limites du calcul des probabilités. (Russian.) *Bulletin de l'Académie Impériale des Sciences* **11** (1917), no. 3, 177–186.

[18] Nicolas Pouyanne: An algebraic approach to Pólya processes. *Ann. Inst. Henri Poincaré Probab. Stat.* **44** (2008), no. 2, 293–323.

Department of Mathematics, Uppsala University, PO Box 480, SE-751 06 Uppsala, Sweden

Email address: `svante.janson@math.uu.se`

URL: `http://www.math.uu.se/svante-janson`