Combinatorics, Probability and Computing (2014) 23, 1148–1186.
doi:10.1017/S0963548314000364

Pólya Urns Via the Contraction Method

MARGARETE KNAPE and RALPH NEININGER

Institute for Mathematics, J. W. Goethe University, 60054 Frankfurt a.M., Germany
(e-mail: {knape,neiningr}@math.uni-frankfurt.de)

Received 15 January 2013; revised 31 October 2013; first published online 1 September 2014

We propose an approach to analysing the asymptotic behaviour of Pólya urns based on the contraction method. For this, a new combinatorial discrete-time embedding of the evolution of the urn into random rooted trees is developed. A decomposition of these trees leads to a system of recursive distributional equations which capture the distributions of the numbers of balls of each colour. Ideas from the contraction method are used to study such systems of recursive distributional equations asymptotically. We apply our approach to a couple of concrete Pólya urns that lead to limit laws with normal limit distributions, with non-normal limit distributions and with asymptotic periodic distributional behaviour.

2010 *Mathematics subject classification*: Primary 60C05
Secondary 60F05, 60J05, 68Q25

1. Introduction

In this paper we develop an approach to proving limit theorems for Pólya urn models by the contraction method. We consider an urn with balls in a finite number $m \geqslant 2$ of different colours, numbered $1, \dots, m$. The evolution of a Pólya urn is determined by an $m \times m$ replacement matrix $R = (a_{ij})_{1 \leqslant i,j \leqslant m}$, which is given in advance together with an initial (time 0) composition of the urn with at least one ball. Time evolves in discrete steps. In each step, one ball is drawn uniformly at random from the urn. If it has colour i it is placed back into the urn together with a_{ij} balls of colour j for all $j = 1, \dots, m$. The steps are iterated independently. A classical problem is to identify the asymptotic behaviour of the numbers of balls of each colour as the number n of steps tends to infinity. The literature on this problem, in particular on limit theorems for the normalized numbers of balls of each colour, is vast. We refer to the monographs of Johnson and Kotz [22] and Mahmoud [26] and the references and comments on the literature in the papers of Janson [16], Flajolet, Gabarró and Pekari [13] and Pouyanne [32].

A couple of approaches have been used to analyse the asymptotic behaviour of Pólya urn models, most notably the method of moments, discrete-time martingale methods, embeddings into continuous-time multitype branching processes, and methods from analytic combinatorics based on generating functions. All these methods use the 'forward' dynamic of the urn process by exploiting the fact that the distribution of the composition at time n given time $n-1$ is explicitly accessible.

In the present paper, we propose an approach based on a 'backward' decomposition of the urn process. We construct a new embedding of the evolution of the urn into an associated combinatorial random tree structure growing in discrete time. Our associated tree can be decomposed at its root (time 0) such that the growth dynamics of the subtrees of the root resemble the whole tree in distribution. More precisely we have different types of distributions for the associated tree, one type for each possible colour of its root. The decomposition of the associated tree into subtrees gives rise to a system of distributional recurrences for the numbers of balls of each colour. To extract the asymptotic behaviour from such systems we develop an approach in the context of the contraction method.

The contraction method is well known in the probabilistic analysis of algorithms. It was introduced by Rösler [34] and first developed systematically in Rachev and Rüschendorf [33]. A rather general framework with numerous applications to the analysis of recursive algorithms and random trees was given by Neininger and Rüschendorf [29]. The contraction method has been used for sequences of distributions of random variables (or random vectors or stochastic processes) that satisfy an appropriate recurrence relation. To the best of our knowledge it has not yet been used for systems of such recurrence relations as they arise in the present paper, the only exception being Leckey, Neininger and Szpankowski [25], where tries are analysed under a Markov source model. A novel technical aspect of the present paper is that we extend the use of the contraction method to systems of recurrence relations systematically.

The aim of this paper is not to compete with other techniques with respect to generality under which urn models can be analysed. Instead we discuss our approach in relation to a couple of examples illustrating the contraction framework in three frequently occurring asymptotic regimes: normal limit laws, non-normal limit laws and regimes with oscillating distributional behaviour. We also discuss the case of random entries in the replacement matrix. Our proofs are generic and can easily be transferred to other urn models or developed into more general theorems when asymptotic expansions of means (respectively means and variances in the normal limit case) are available: see the types of expansions of the means in Section 3.

A general assumption in the present paper is that the replacement matrix is balanced, *i.e.*, we have $\sum_{j=1}^{m} a_{ij} =: K-1$ for all $i = 1,\dots,m$, where $K \geqslant 2$ is a fixed integer. (The notation K is unfortunate since this integer is not random, and it has mainly been chosen because of similarity in notation to earlier work on the contraction method.) An implication of the balance condition is that the asymptotic growth of the subtrees of the associated tree processes can jointly be captured by Dirichlet distributions. This leads to characterizations of the limit distributions in all cases (normal, non-normal and oscillatory behaviour) by systems (see (3.2)–(3.6) below) of distributional fixed point equations where all coefficients are powers of components of a Dirichlet-distributed vector; see also the discussion in Section 3. The present approach reveals that all three regimes are governed by systems of distributional fixed point equations of similar type.

The paper is organized as follows. In Section 2 we introduce the associated trees into which the urn models are embedded and derive the systems of distributional recurrences for the numbers of balls of a certain colour from the associated trees. In Section 3 we outline the types of systems of fixed point equations that emerge from the distributional recurrences after proper normalization. To make these recurrences and fixed point equations accessible to the contraction method, in Section 4 we first introduce spaces of probability distributions and appropriate Cartesian product spaces together with metrics on these product spaces. The metrics in use are product versions of the minimal L_p-metrics and product versions of the Zolotarev metrics. In Section 5 we use these spaces and metrics to show that our systems of distributional fixed point equations uniquely characterize vectors of probability distributions via a contraction property. These cover the types of distributional fixed point equations that appear in the final Section 6, where we discuss examples of limit laws for Pólya urn schemes within our approach. Also in Section 6, our convergence proofs are worked out, again based on the product versions of the minimal L_p and Zolotarev metrics. In Section 7 we compare our study of systems of recurrences with an alternative formulation based on multivariate recurrences and explain the advantages and necessity of our approach.

For similar results see [9] (announced after posting the present paper on `arXiv.org`).

Notation. We let $\xrightarrow{d}$ denote convergence in distribution, and we let $\mathcal{N}(\mu, \sigma^2)$ denote the normal distribution on $\mathbb{R}$ with mean $\mu \in \mathbb{R}$ and variance $\sigma^2 \geqslant 0$. In the case $\sigma^2 = 0$, this degenerates to the Dirac measure in μ. Throughout the paper, Bachmann–Landau symbols are used in asymptotic statements. We let $\log(x)$ for $x > 0$ be the natural logarithm of x and denote the non-negative integers by $\mathbb{N}_0 := \{0, 1, 2, \ldots\}$.

2. A recursive description of Pólya urns

In this section we explain our embedding of urn processes into associated combinatorial random tree structures growing in discrete time. The distributional self-similarity within the subtrees of the roots of these associated trees leads to systems of distributional recurrences which constitute the core of our approach.

The Pólya urn. To develop our approach, we first consider an urn model with two colours, black and white, and a deterministic replacement matrix R. Below, an extension of this approach to urns with more than two colours and replacement matrices with random entries is discussed too. To be definite, we use the replacement matrix

$$R = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \quad \text{with } a, d \in \mathbb{N}_0 \cup \{-1\} \text{ and } b, c \in \mathbb{N}_0, \tag{2.1}$$

with

$$a + b = c + d =: K - 1 \geqslant 1.$$

The assumption that the sums of the entries in each row are the same will become essential only from Lemma 2.1 on. Now, after drawing a black ball, this ball is placed back into the urn together with a new black balls and b new white balls. If a white ball is drawn, it

https://doi.org/10.1017/S0963548314000364 Published online by Cambridge University Press

is placed back into the urn together with c black balls and d white balls. A diagonal entry $a = -1$ (or $d = -1$) implies that a drawn black (or white) ball is not placed back into the urn while balls of the other colour are still added to the urn. As initial configuration, we consider both one black ball and one white ball. Other initial configurations can be dealt with as well, also discussed below. We let B_n^{b} denote the number of black balls after n steps when initially starting with one black ball, and we let B_n^{w} denote the number of black balls after n steps when initially starting with one white ball. Hence, we have $B_0^{\mathrm{b}} = 1$ and $B_0^{\mathrm{w}} = 0$.

The associated tree. We encode the urn process as follows by a discrete-time evolution of a random tree with nodes coloured black or white. This tree is called an *associated tree*. The initial urn with one ball, say a black one, is associated with a tree with one root node of the same (black) colour. The ball in the urn is represented by this root node. Now drawing the ball and placing it back into the urn together with a new black balls and b new white balls is encoded in the associated tree by adding $a + b + 1 = K$ children to the root node, $a + 1$ of them being black and b being white. The root node then no longer represents a ball in the tree, whereas the K new leaves of the tree now represent the K balls in the urn. Now, we iterate this procedure. At any step, a ball is drawn from the urn. It is represented by one of the leaves, say node v in the tree. The urn follows its dynamic. If the ball drawn is black, the (black) leaf v gets K children, $a + 1$ black ones and b white ones. Similarly, if the ball drawn is white, the (white) leaf v gets c black children and $d + 1$ white children. In both cases, v no longer represents a ball in the urn. The ball drawn and the new balls are represented by the children of v. The correspondence between all other leaves of the tree and the other balls in the urn remains unchanged. For an example of an evolution of an urn and its associated tree, see Figure 1. Hence, at any time, the balls in the urn are represented by the leaves of the associated tree, where the colours of balls and representing leaves match. Each node of the tree is either a leaf or has K children. We could also simulate the urn process by only running the evolution of the associated tree as follows. Start with one root node of the colour of the initial ball of the urn. At any step, choose one of the leaves of the tree uniformly at random, inspect its colour, add K children to the chosen leaf and colour these children as defined above. Then, after n steps, the tree has $n(K-1)+1$ leaves. The number of black leaves is distributed as B_n^{b} if the root node was black, and as B_n^{w} if the root node was white.

Subsequently, it is important to note the following recursive structure of the associated tree. For a fixed replacement matrix of the Pólya urn, we consider the two initial compositions of one black ball, respectively one white ball, and their two associated trees. We call these the b-associated, respectively w-associated tree. Consider one of these associated trees after $n \geqslant 1$ steps. It has $n(K-1)+1$ leaves, and each subtree rooted at a child of the associated tree's root (we call them subtrees for short) has a random number of leaves according to how often a leaf node has been chosen for replacement in the subtree. We condition on the numbers of leaves of the subtrees being $i_r(K-1)+1$ with $i_r \in \mathbb{N}_0$ for $r = 1, \ldots, K$. Note that we have $\sum_{r=1}^{K} i_r = n - 1$, the -1 resulting from the fact that in the first step of the evolution of the associated tree, the subtrees are being generated; only afterwards do they start growing. From the evolution of the b-associated tree, it is clear

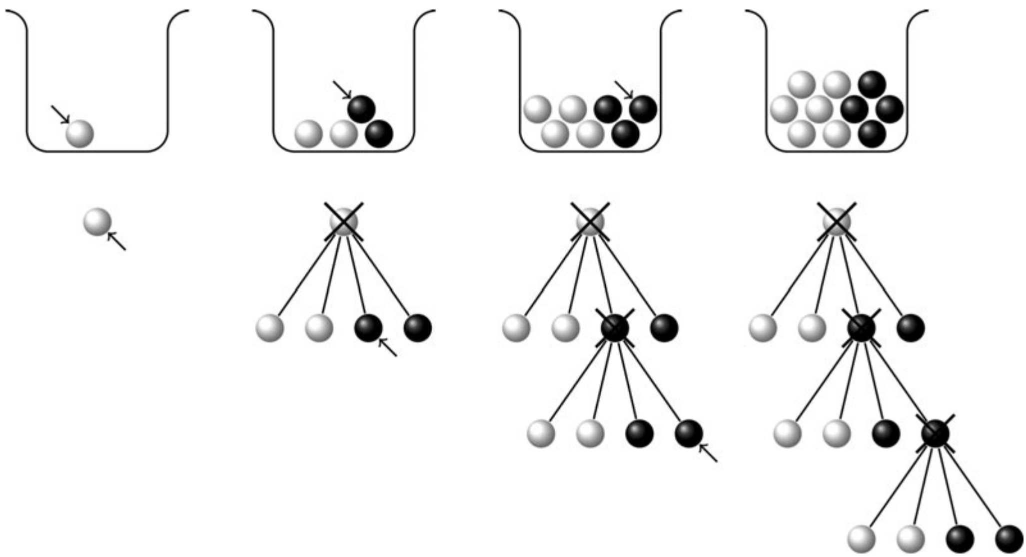

Figure 1. A realization of the evolution of the Pólya urn with replacement matrix $\left[\begin{smallmatrix}1 & 2\\ 2 & 1\end{smallmatrix}\right]$ and initially one white ball. The arrows indicate which ball is drawn (resp. which leaf is replaced) in each step. The associated tree is shown below each urn. Leaf nodes correspond to the balls in the urn; non-leaf nodes (crossed out) no longer correspond to balls in the urn. However, their colour still matters for the recursive decomposition of the associated tree.

that, conditioned on the subtrees' numbers of leaves being $i_r(K-1)+1$, the subtrees are stochastically independent and the rth subtree is distributed as an associated tree after i_r steps. Whether it has the distribution of the b-associated tree or the w-associated tree depends on the colour of the subtree's root node.

To summarize, we have that conditioned on their numbers of leaves, the subtrees of associated trees are independent and distributed as associated trees of corresponding size and type inherited from the colour of their root node.

System of recursive equations. We set up recursive equations for the distributions of the quantities B_n^{b} and B_n^{w}. For B_n^{b}, we start the urn with one black ball and get a b-associated tree with a black root node. Now, B_n^{b} is distributed as the number of black leaves in the associated tree after n steps which, for $n \geqslant 1$, we express as the sum of the numbers of black leaves of its subtrees. As discussed above, conditionally on $I^{(n)} = (I_1^{(n)}, \dots, I_K^{(n)})$, the vector of the numbers of balls drawn in each subtree, these subtrees are independent and distributed as b-associated trees or w-associated trees of the corresponding size depending on the colour of their roots. In a b-associated tree, the root has $a+1$ black and $b = K-(a+1)$ white children. Hence, we obtain

$$B_n^{\mathrm{b}} \stackrel{d}{=} \sum_{r=1}^{a+1} B_{I_r^{(n)}}^{\mathrm{b},(r)} + \sum_{r=a+2}^{K} B_{I_r^{(n)}}^{\mathrm{w},(r)}, \quad n \geqslant 1, \tag{2.2}$$

where $\stackrel{d}{=}$ denotes that the left- and right-hand sides have an identical distribution; we have that $(B_k^{\mathrm{b},(1)})_{0\leqslant k<n}, \dots, (B_k^{\mathrm{b},(a+1)})_{0\leqslant k<n}$, $(B_k^{\mathrm{w},(a+2)})_{0\leqslant k<n}, \dots, (B_k^{\mathrm{w},(K)})_{0\leqslant k<n}$, $I^{(n)}$ are independent,

https://doi.org/10.1017/S0963548314000364 Published online by Cambridge University Press

the $B_k^{\mathrm{b},(r)}$ are distributed as B_k^{b}, the $B_k^{\mathrm{w},(r)}$ are distributed as B_k^{w} for $k=0,\dots,n-1$ for the respective values of r.

Similarly, we obtain a recursive distributional equation for B_n^{w}. We have

$$B_n^{\mathrm{w}} \stackrel{d}{=} \sum_{r=1}^{c} B_{I_r^{(n)}}^{\mathrm{b},(r)} + \sum_{r=c+1}^{K} B_{I_r^{(n)}}^{\mathrm{w},(r)}, \quad n \geqslant 1, \tag{2.3}$$

with conditions on independence and identical distributions as in (2.2). Note that with the initial value $(B_0^{\mathrm{b}}, B_0^{\mathrm{w}}) = (1,0)$, the system of equations (2.2)–(2.3) defines the sequence of pairs of distributions $(\mathcal{L}(B_n^{\mathrm{b}}), \mathcal{L}(B_n^{\mathrm{w}}))_{n\geqslant 0}$.

General number of colours. The approach above for urns with two colours extends directly to urns with an arbitrary number $m \geqslant 2$ of colours. We denote the replacement matrix by $R=(a_{ij})_{1\leqslant i,j\leqslant m}$ with

$$a_{ij} \in \begin{cases} \mathbb{N}_0 & \text{for } i \neq j, \\ \mathbb{N}_0 \cup \{-1\} & \text{for } i=j, \end{cases} \quad \text{and} \quad \sum_{j=1}^{m} a_{ij} =: K-1 \geqslant 1 \quad \text{for } i=1,\dots,m.$$

The colours (subsequently also called types) are now numbered $1,\dots,m$ and we focus on the number of balls of type 1 after n steps. When starting with one ball of type j we let $B_n^{[j]}$ denote the number of type 1 balls after n steps. To formulate a system of distributional recurrences generalizing (2.2) and (2.3), we further denote the intervals of integers:

$$J_{ij} := \begin{cases} \left[1+\sum_{k<i} a_{kj}, \sum_{k\leqslant i} a_{kj}\right] \cap \mathbb{N}_0 & \text{for } i<j, \\ \left[1+\sum_{k<i} a_{kj}, 1+\sum_{k\leqslant i} a_{kj}\right] \cap \mathbb{N}_0 & \text{for } i=j, \\ \left[2+\sum_{k<i} a_{kj}, 1+\sum_{k\leqslant i} a_{kj}\right] \cap \mathbb{N}_0 & \text{for } i>j, \end{cases} \tag{2.4}$$

with the convention $[x,y]=\emptyset$ if $x>y$. Then, we have

$$B_n^{[j]} \stackrel{d}{=} \sum_{i=1}^{m} \sum_{r\in J_{ij}} B_{I_r^{(n)}}^{[i],(r)}, \quad n\geqslant 1, \quad j\in\{1,\dots,m\}, \tag{2.5}$$

where, for each $j\in\{1,\dots,m\}$, we have that the family

$$\left\{\left(B_k^{[i],(r)}\right)_{0\leqslant k<n} \mid r\in J_{ij}, i\in\{1,\dots,m\}\right\} \cup \{I^{(n)}\}$$

is independent, $B_k^{[i],(r)}$ is distributed as $B_k^{[i]}$ for all $i\in\{1,\dots,m\}$, $0\leqslant k<n$ and $r\in J_{ij}$ and $I^{(n)}$ has the distribution as above in Lemma 2.1.

Composition vectors. For urns with more than two colours one may study the numbers of balls of each colour jointly. Even though the system (2.5) only gives access to the marginals of this composition vector, we could also derive a system of recurrences for the composition vectors and develop our approach for the joint distribution of the composition vector. The work spaces $(\mathcal{M}_s^{\mathbb{R}})^{\times d}$ and $(\mathcal{M}_s^{\mathbb{C}})^{\times d}$ defined in Section 4 below (there d corresponds to the number of colours) then become $(\mathcal{M}_s^{\mathbb{R}^{d-1}})^{\times d}$ and $(\mathcal{M}_s^{\mathbb{C}^{d-1}})^{\times d}$. The Zolotarev metrics ζ_s and minimal L_p-metrics ℓ_p are defined on $\mathbb{R}^{d-1}$ and $\mathbb{C}^{d-1}$ as well

https://doi.org/10.1017/S0963548314000364 Published online by Cambridge University Press

and can be used to develop a similar limit theory for the composition vectors as presented here for their marginals.

Random entries in the replacement matrix. The case of a replacement matrix with random entries such that each row almost surely sums to a deterministic and fixed $K-1 \geqslant 1$ can be covered by an extension of the system (2.5). Instead of formulating such an extension explicitly, we discuss an example in Section 6.2.

Growth of subtrees. In our analysis, the asymptotic growth of the K subtrees of the associated tree is used. We denote by $I^{(n)} = (I_1^{(n)}, \ldots, I_K^{(n)})$ the vector of the numbers of draws of leaves from each subtree after $n \geqslant 1$ draws in the full associated tree. In other words, $I_r^{(n)}(K-1)+1$ is the number of leaves of the rth subtree after $n \geqslant 1$ steps. We have $I^{(1)} = (0, \ldots, 0)$, and $I^{(2)}$ is a vector with all entries being 0, except for one coordinate which is 1. To describe the asymptotic growth of $I^{(n)}$, we need the Dirichlet distribution Dirichlet$((K-1)^{-1}, \ldots, (K-1)^{-1})$: it is the distribution of a random vector $(D_1, \ldots, D_K)$ with $\sum_{r=1}^K D_r = 1$ and such that $(D_1, \ldots, D_{K-1})$ has a Lebesgue density supported by the simplex

$$\mathcal{S}_K := \left\{ (x_1, \ldots, x_{K-1}) \in [0,1]^{K-1} \;\middle|\; \sum_{r=1}^{K-1} x_r \leqslant 1 \right\}$$

given for $x \in \mathcal{S}_K$ by

$$x = (x_1, \ldots, x_{K-1}) \mapsto c_K \left(1 - \sum_{r=1}^{K-1} x_r\right)^{\frac{2-K}{K-1}} \prod_{r=1}^{K-1} x_r^{\frac{2-K}{K-1}}, \quad c_K = \frac{\Gamma\big((K-1)^{-1}\big)^{1-K}}{K-1},$$

where Γ denotes Euler's gamma function. In particular, $D_1, \ldots, D_K$ are identically distributed with the beta$\big((K-1)^{-1}, 1\big)$ distribution, *i.e.*, with Lebesgue density

$$x \mapsto (K-1)^{-1} x^{\frac{2-K}{K-1}}, \quad x \in [0,1].$$

We have the following asymptotic behaviour of $I^{(n)}$.

Lemma 2.1. *Consider a Pólya urn with constant row sum $K-1 \geqslant 1$ and its associated tree. For the numbers of balls $I^{(n)} = (I_1^{(n)}, \ldots, I_K^{(n)})$ drawn in each subtree of the associated tree when n balls have been drawn in the whole associated tree, we have, as $n \to \infty$,*

$$\left(\frac{I_1^{(n)}}{n}, \ldots, \frac{I_K^{(n)}}{n}\right) \longrightarrow (D_1, \ldots, D_K)$$

almost surely and in any L_p, where $(D_1, \ldots, D_K)$ has the Dirichlet distribution

$$\mathcal{L}(D_1, \ldots, D_K) = \text{Dirichlet}\left(\frac{1}{K-1}, \ldots, \frac{1}{K-1}\right).$$

Proof. The sequence $(I_1^{(n)}(K-1)+1, \ldots, I_K^{(n)}(K-1)+1)_{n \in \mathbb{N}_0}$ has an interpretation by another urn model, which we call the subtree-induced urn. For this, we give additional labels to the leaves of the associated tree. The set of possible labels is $\{1, \ldots, K\}$, and we label a leaf j if it belongs to the jth subtree of the root (any ordering of the subtrees of

https://doi.org/10.1017/S0963548314000364 Published online by Cambridge University Press

the root is fine). Hence, all leaves of a subtree of the associated tree's root get the same label, and leaves of different subtrees get different labels. Now, the subtree-induced urn has balls of colours $1,\dots,K$. At any time, the number of balls of each colour is identical to the numbers of leaves with the corresponding label. Hence, the dynamic of the subtree-induced urn is that of a Pólya urn with initially K balls, one of each colour. Whenever a ball is drawn, it is placed back into the urn together with $K-1$ balls of the same colour. In other words, the replacement matrix for the dynamic of the subtree-induced urn is a $K \times K$ diagonal matrix with all diagonal entries equal to $K-1$. After n steps, we have $I_r^{(n)}(K-1)+1$ balls of colour r. The dynamic of the subtree-induced urn as a K-colour Pólya–Eggenberger urn is well known (see Athreya [1, Corollary 1]): for $n \to \infty$, almost surely and in L_p for any $p \geqslant 1$, we have

$$\left(\frac{I_1^{(n)}(K-1)+1}{n(K-1)+1},\dots,\frac{I_K^{(n)}(K-1)+1}{n(K-1)+1}\right) \longrightarrow (D_1,\dots,D_K),$$

where $(D_1,\dots,D_K)$ has a Dirichlet$((K-1)^{-1},\dots,(K-1)^{-1})$ distribution. This implies the assertion. □

Subsequently we only consider balanced urns such that we have the asymptotic behaviour of $I^{(n)}/n$ in Lemma 2.1 available. The assumption of balance only enters our subsequent analysis via Lemma 2.1. It also seems feasible to apply our approach to unbalanced urns that have an associated tree such that $I^{(n)}/n$ converges to a non-degenerate limit vector $V=(V_1,\dots,V_K)$ of random probabilities, *i.e.*, of random $V_1,\dots,V_K \geqslant 0$ such that $\sum_{r=1}^K V_r = 1$ almost surely and $\mathbb{P}(\max_{1\leqslant r\leqslant K} V_r < 1) > 0$. It seems that the contraction argument may even allow the distribution of V to depend on the initial colour of the ball in the urn. We leave these issues for future research.

3. Systems of limit equations

In this section we outline how systems of the form (2.5) are used subsequently. Based on the order of means and variances, the $B_n^{[j]}$ are normalized and recurrences for the normalized random variables are considered. From this, with $n \to \infty$, we derive systems of recursive distributional equations; see (3.2), (3.4) and (3.6). According to the general idea of the contraction method, we then show first that these systems characterize distributions (see Section 5), and second that the normalized random variables converge in distribution towards these distributions (see Section 6). In the periodic case (c) we do not have convergence, but the solution of system (3.6) allows us to describe the asymptotic periodic behaviour.

Particularly crucial are the expansions of the means

$$\mu_n^{[j]} := \mathbb{E}\big[B_n^{[j]}\big], \quad j=1,\dots,m,$$

which are intimately related to the spectral decomposition of the replacement matrix. We only consider cases where these means grow linearly. Note, however, that even balanced urns can have quite different growth orders. An example is the replacement matrix $\left[\begin{smallmatrix}4&0\\3&1\end{smallmatrix}\right]$;

https://doi.org/10.1017/S0963548314000364 Published online by Cambridge University Press

see Kotz, Mahmoud and Robert [24] for this example or Janson [17] for a comprehensive account of urns with triangular replacement matrix.

Type (a). Assume that we have expansions of the form, as $n \to \infty$,

$$\mu_n^{[j]} = c_\mu n + d_j n^\lambda + o(n^\lambda), \quad j = 1, \dots, m,$$

with a constant $c_\mu > 0$ independent of j, with constants $d_j \in \mathbb{R}$ and an exponent $1/2 < \lambda < 1$. We call this scenario type (a). This suggests that the variances are of the order $n^{2\lambda}$ and a proper scaling is

$$X_n^{[j]} := \frac{B_n^{[j]} - \mu_n^{[j]}}{n^\lambda}, \quad n \geqslant 1, \quad j = 1, \dots, m. \tag{3.1}$$

Deriving from (2.5) a system of recurrences for the $X_n^{[j]}$ and letting formally $n \to \infty$ (this is done explicitly in the examples in Section 6), we obtain the system of fixed point equations

$$X^{[j]} \stackrel{d}{=} \sum_{i=1}^{m} \sum_{r \in J_{ij}} D_r^\lambda X^{[i],(r)} + b^{[j]}, \quad j = 1, \dots, m, \tag{3.2}$$

where the $X^{[i],(r)}$ and the $(D_1, \dots, D_K)$ are independent, $X^{[i],(r)}$ are distributed as $X^{[i]}$, the $(D_1, \dots, D_K)$ is distributed as in Lemma 2.1 and the $b^{[j]}$ are functions of $(D_1, \dots, D_K)$. It turns out that such a system subject to centred $X^{[j]}$ with finite second moments has a unique solution on the level of distributions (Theorem 5.1). This identifies the weak limits of the $X_n^{[j]}$. Examples are given in Sections 6.1 and 6.2. One can also obtain the same system (3.2) with $b^{[j]} = 0$ for all j by only centering the $B_n^{[j]}$ by $c_\mu n$ instead of the exact mean. Then system (3.2) has to be solved subject to finite second moments and appropriate means. Moreover, the system allows us to calculate higher-order moments of the solution. From the second and third moments one can typically see that the solution is not a vector of normal distributions.

Expansions of the form

$$\mu_n^{[j]} = c_\mu n + d_j n^\lambda \log^\nu(n) + o(n^\lambda \log^\nu(n)), \quad j = 1, \dots, m,$$

with $\nu \geqslant 1$, also appear; see Janson [16] or the table on page 279 of Pouyanne [31] for a classification. Such additional factors $\log^\nu(n)$, slowly varying at infinity, give rise to the same limit system (3.2) and hence do not affect the limit distributions. These cases can be covered in a similar way to the examples in Section 6. We omit the details; see, however, Hwang and Neininger [14] for the occurrence and analysis of similar slowly varying factors.

Type (b). Assume that we have expansions of the form, as $n \to \infty$,

$$\mu_n^{[j]} = c_\mu n + o(\sqrt{n}), \quad j = 1, \dots, m,$$

with a constant $c_\mu > 0$ independent of j. We call this scenario type (b). This suggests that the variances are of linear order and a proper scaling is

$$X_n^{[j]} := \frac{B_n^{[j]} - \mu_n^{[j]}}{\sqrt{\mathrm{Var}(B_n^{[j]})}}, \quad n \geqslant 1, \quad j = 1, \dots, m \tag{3.3}$$

https://doi.org/10.1017/S0963548314000364 Published online by Cambridge University Press

(or $\sqrt{\mathrm{Var}(B_n^{[j]})}$ replaced by $\sqrt{n}$). The corresponding system of fixed point equations in the limit is

$$X^{[j]} \stackrel{d}{=} \sum_{i=1}^{m} \sum_{r \in J_{ij}} \sqrt{D_r} X^{[i],(r)}, \quad j = 1, \ldots, m, \tag{3.4}$$

with conditions as in (3.2). Under appropriate assumptions on moments we find that the only solution is for all $X^{[j]}$ to be standard normally distributed (Theorem 5.2). This leads to asymptotic normality of the $X_n^{[j]}$. Examples are given in Sections 6.1 and 6.2. The case

$$\mu_n^{[j]} = c_\mu n + \Theta(\sqrt{n}), \quad j = 1, \ldots, m,$$

leads to the same system of fixed point equations (3.4). However, here the variances are typically of order $n \log^{\delta}(n)$ with a positive δ.

Type (c). Assume that we have expansions of the form, as $n \to \infty$,

$$\mu_n^{[j]} = c_\mu n + \Re\big(\kappa_j n^{\mathrm{i}\mu}\big) n^{\lambda} + o(n^{\lambda}), \quad j = 1, \ldots, m,$$

with a constant $c_\mu > 0$ independent of j, $1/2 < \lambda < 1$, constants $\kappa_j \in \mathbb{C}$ and $\mu \in \mathbb{R} \setminus \{0\}$ (where i denotes the imaginary unit). We call this scenario type (c). This suggests oscillating variances of order $n^{2\lambda}$. The oscillatory behaviour of mean and variance typically cannot be removed by proper scaling to obtain convergence towards a limit distribution. Using the scaling

$$X_n^{[j]} := \frac{B_n^{[j]} - c_\mu n}{n^{\lambda}}, \quad n \geqslant 1, \quad j = 1, \ldots, m, \tag{3.5}$$

it turns out that the oscillating behaviour of the $X_n^{[j]}$ can be captured by the system of fixed point equations

$$X^{[j]} \stackrel{d}{=} \sum_{i=1}^{m} \sum_{r \in J_{ij}} D_r^{\omega} X^{[i],(r)}, \quad j = 1, \ldots, m, \tag{3.6}$$

with conditions as in (3.2) and $\omega := \lambda + \mathrm{i}\mu$. Under appropriate moment assumptions this has a unique solution within distributions on $\mathbb{C}$ (Theorem 5.3). An example of a corresponding distributional approximation is given in Section 6.3.

As in type (a) we may have additional factors $\log^{\nu}(n)$, *i.e.*,

$$\mu_n^{[j]} = c_\mu n + \Re(\kappa_j n^{\mathrm{i}\mu}) n^{\lambda} \log^{\nu}(n) + o(n^{\lambda} \log^{\nu}(n)), \quad j = 1, \ldots, m.$$

The comments for type (a) cases above apply here as well.

Note that the approach of embedding urn models into continuous-time multitype branching processes (see [2, 16]) also leads to characterizations of the limit distributions as in (3.2) and (3.6). However, the form of the fixed point equations is different; see the system in equation (3.5) in Janson [16]. Properties of such fixed points have been studied by Chauvin, Pouyanne and Sahnoun [10, 8, 7].

4. Spaces of distributions and metrics

In this section we define Cartesian products of spaces of probability distributions and metrics on these products. These metric spaces will be used below, first to characterize limit distributions of urn models (Section 5) and then to prove convergence in distribution of the scaled numbers of balls of a colour (Section 6).

Spaces. We let $\mathcal{M}^{\mathbb{R}}$ denote the space of all probability distributions on $\mathbb{R}$ with the Borel σ-field. Moreover, we consider the subspaces

$$\begin{aligned}\mathcal{M}_s^{\mathbb{R}} &:= \{\mathcal{L}(X) \in \mathcal{M}^{\mathbb{R}} \mid \mathbb{E}[|X|^s] < \infty\}, \quad s > 0,\\ \mathcal{M}_s^{\mathbb{R}}(\mu) &:= \{\mathcal{L}(X) \in \mathcal{M}_s^{\mathbb{R}} \mid \mathbb{E}[X] = \mu\}, \quad s \geqslant 1, \mu \in \mathbb{R},\\ \mathcal{M}_s^{\mathbb{R}}(\mu, \sigma^2) &:= \{\mathcal{L}(X) \in \mathcal{M}_s^{\mathbb{R}}(\mu) \mid \mathrm{Var}(X) = \sigma^2\}, \quad s \geqslant 2, \mu \in \mathbb{R}, \sigma \geqslant 0.\end{aligned}$$

We need the d-fold Cartesian products, $d \in \mathbb{N}$, of these spaces denoted by

$$(\mathcal{M}_s^{\mathbb{R}})^{\times d} := \mathcal{M}_s^{\mathbb{R}} \times \cdots \times \mathcal{M}_s^{\mathbb{R}}, \tag{4.1}$$

and analogously $(\mathcal{M}_s^{\mathbb{R}}(\mu))^{\times d}$ and $(\mathcal{M}_s^{\mathbb{R}}(\mu, \sigma^2))^{\times d}$.

We also need probability distributions on the complex plane $\mathbb{C}$. We let $\mathcal{M}^{\mathbb{C}}$ denote the space of all probability distributions on $\mathbb{C}$ with the Borel σ-field. Moreover, for $\gamma \in \mathbb{C}$ we use the subspaces and product space

$$\begin{aligned}\mathcal{M}_s^{\mathbb{C}} &:= \{\mathcal{L}(X) \in \mathcal{M}^{\mathbb{C}} \mid \mathbb{E}[|X|^s] < \infty\}, \quad s > 0,\\ \mathcal{M}_2^{\mathbb{C}}(\gamma) &:= \{\mathcal{L}(X) \in \mathcal{M}_2^{\mathbb{C}} \mid \mathbb{E}[X] = \gamma\},\\ (\mathcal{M}_2^{\mathbb{C}}(\gamma))^{\times d} &:= \mathcal{M}_2^{\mathbb{C}}(\gamma) \times \cdots \times \mathcal{M}_2^{\mathbb{C}}(\gamma).\end{aligned}$$

To cover the different behaviour of the urns, two types of metrics are constructed: extensions of the Zolotarev metrics ζ_s and the minimal L_p-metric ℓ_p to the product spaces defined above.

Zolotarev metric. The Zolotarev metric was introduced and studied in [39, 40]. The contraction method based on the Zolotarev metric was systematically developed in [29] and, for issues that go beyond what is needed in this paper, in [20] and [30]. We only need the following properties. For distributions $\mathcal{L}(X), \mathcal{L}(Y) \in \mathcal{M}^{\mathbb{R}}$ the Zolotarev distance ζ_s, $s > 0$, is defined by

$$\zeta_s(X, Y) := \zeta_s(\mathcal{L}(X), \mathcal{L}(Y)) := \sup_{f \in \mathcal{F}_s} |\mathbb{E}[f(X) - f(Y)]|, \tag{4.2}$$

where $s = m + \alpha$ with $0 < \alpha \leqslant 1$, $m \in \mathbb{N}_0$, and

$$\mathcal{F}_s := \{f \in C^m(\mathbb{R}, \mathbb{R}) : |f^{(m)}(x) - f^{(m)}(y)| \leqslant |x - y|^\alpha\}, \tag{4.3}$$

the space of m-times continuously differentiable functions from $\mathbb{R}$ to $\mathbb{R}$ such that the mth derivative is Hölder-continuous of order α with Hölder constant 1.

We have that $\zeta_s(X, Y) < \infty$ if all moments of orders $1, \dots, m$ of X and Y are equal and if the sth absolute moments of X and Y are finite. Since the cases $1 < s \leqslant 3$ are used later on, we have two basic cases. First, for $1 < s \leqslant 2$ we have $\zeta_s(X, Y) < \infty$ for $\mathcal{L}(X)$, $\mathcal{L}(Y) \in \mathcal{M}_s^{\mathbb{R}}(\mu)$ for any $\mu \in \mathbb{R}$. Second, for $2 < s \leqslant 3$ we have $\zeta_s(X, Y) < \infty$ for $\mathcal{L}(X)$,

$\mathcal{L}(Y) \in \mathcal{M}_s^{\mathbb{R}}(\mu, \sigma^2)$ for any $\mu \in \mathbb{R}$ and $\sigma \geqslant 0$. Moreover, the pairs $(\mathcal{M}_s^{\mathbb{R}}(\mu), \zeta_s)$ for $1 < s \leqslant 2$ and $(\mathcal{M}_s^{\mathbb{R}}(\mu, \sigma^2), \zeta_s)$ for $2 < s \leqslant 3$ are complete metric spaces; for completeness see [11, Theorem 5.1].

Convergence in ζ_s implies weak convergence on $\mathbb{R}$. Furthermore, ζ_s is $(s,+)$-ideal, *i.e.*, we have

$$\zeta_s(X+Z, Y+Z) \leqslant \zeta_s(X,Y), \quad \zeta_s(cX, cY) = c^s \zeta_s(X,Y) \tag{4.4}$$

for all Z independent of (X,Y) and all $c > 0$. Note that this implies that, for $X_1, \dots, X_n$ independent and $Y_1, \dots, Y_n$ independent such that the respective ζ_s distances are finite, we have

$$\zeta_s\left(\sum_{i=1}^n X_i, \sum_{i=1}^n Y_i\right) \leqslant \sum_{i=1}^n \zeta_s(X_i, Y_i). \tag{4.5}$$

On the product spaces $(\mathcal{M}_s^{\mathbb{R}}(\mu))^{\times d}$ for $1 < s \leqslant 2$ and $(\mathcal{M}_s^{\mathbb{R}}(\mu, \sigma^2))^{\times d}$ for $2 < s \leqslant 3$, our first main tool is

$$\zeta_s^{\vee}((\nu_1, \dots, \nu_d), (\mu_1, \dots, \mu_d)) := \max_{1 \leqslant j \leqslant d} \zeta_s(\nu_j, \mu_j),$$

where $(\nu_1, \dots, \nu_d), (\mu_1, \dots, \mu_d) \in \mathcal{M}_s^{\mathbb{R}}(\mu))^{\times d}$ and $\in (\mathcal{M}_s^{\mathbb{R}}(\mu, \sigma^2))^{\times d}$ respectively. Note that $\zeta_s^{\vee}$ is a complete metric on the respective product spaces and induces the product topology.

Minimal L_p-metric ℓ_p. First, for probability metrics on the real line, the minimal L_p-metric ℓ_p, $1 \leqslant p < \infty$ is defined by

$$\ell_p(\nu, \varrho) := \inf\{\|V - W\|_p \mid \mathcal{L}(V) = \nu, \mathcal{L}(W) = \varrho\}, \quad \nu, \varrho \in \mathcal{M}_p^{\mathbb{R}},$$

where

$$\|V - W\|_p := (\mathbb{E}[|V - W|^p])^{1/p}$$

is the usual L_p-norm. The spaces $(\mathcal{M}_p^{\mathbb{R}}, \ell_p)$ and $(\mathcal{M}_p^{\mathbb{R}}(\mu), \ell_p)$ for $1 \leqslant p < \infty$ are complete metric spaces: see [6]. The infimum in the definition of ℓ_p is a minimum. Random variables V', W', with distributions ν and ϱ, respectively, such that $\ell_p(\nu, \varrho) = \|V' - W'\|_p$ are called *optimal couplings*. They exist for all $\nu, \varrho \in \mathcal{M}_p^{\mathbb{R}}$. We use the notation $\ell_p(X,Y) := \ell_p(\mathcal{L}(X), \mathcal{L}(Y))$ for random variables X and Y. Subsequently the following inequality between the ℓ_p- and ζ_s-metrics is used:

$$\zeta_s(X,Y) \leqslant \left((\mathbb{E}[|X|^s])^{1-1/s} + (\mathbb{E}[|Y|^s])^{1-1/s}\right)\ell_s(X,Y), \quad 1 < s \leqslant 3, \tag{4.6}$$

where for $1 < s \leqslant 2$ we need $\mathcal{L}(X), \mathcal{L}(Y) \in \mathcal{M}_s^{\mathbb{R}}(\mu)$ for some $\mu \in \mathbb{R}$, and for $2 < s \leqslant 3$ we need $\mathcal{L}(X), \mathcal{L}(Y) \in \mathcal{M}_s^{\mathbb{R}}(\mu, \sigma^2)$ for some $\mu \in \mathbb{R}$ and $\sigma \geqslant 0$ (see [11, Lemma 5.7]).

On the product space $(\mathcal{M}_2^{\mathbb{R}}(0))^{\times d}$, we define

$$\ell_2^{\vee}((\nu_1, \dots, \nu_d), (\varrho_1, \dots, \varrho_d)) := \max_{1 \leqslant j \leqslant d} \ell_2(\nu_j, \varrho_j),$$

where $(\nu_1, \dots, \nu_d), (\mu_1, \dots, \mu_d) \in (\mathcal{M}_2^{\mathbb{R}}(0))^{\times d}$. Note that $(\mathcal{M}_2^{\mathbb{R}}(0))^{\times d}, \ell_2^{\vee})$ is a complete metric space as well.

Second, on the complex plane the minimal L_p-metric ℓ_p is defined similarly by

$$\ell_p(\nu, \varrho) := \inf\{\|V - W\|_p \mid \mathcal{L}(V) = \nu, \mathcal{L}(W) = \varrho\}, \quad \nu, \varrho \in \mathcal{M}_p^{\mathbb{C}},$$

with the analogous definition of the L_p-norm. The respective metric spaces are complete as in the real case and optimal couplings exist as well. On the product space $(\mathcal{M}_2^{\mathbb{C}}(0))^{\times d}$ we use

$$\ell_2^{\vee}((v_1,\dots,v_d),(\varrho_1,\dots,\varrho_d)) := \max_{1\leqslant j\leqslant d} \ell_2(v_j,\varrho_j),$$

where $(v_1,\dots,v_d),(\mu_1,\dots,\mu_d) \in (\mathcal{M}_2^{\mathbb{C}}(0))^{\times d}$. Note that $((\mathcal{M}_2^{\mathbb{C}}(0))^{\times d},\ell_2^{\vee})$ is a complete metric space as well.

Preview of the use of spaces and metrics. The guidance as to which space and metric to use in which asymptotic regime of Pólya urns is as follows. We return to the three types (a)–(c) of urns from the previous section.

(a) Urns that, after scaling, lead to convergence to a non-normal limit distribution. Typically such a convergence holds almost surely, but we only discuss convergence in distribution.
(b) Urns that, after scaling, lead to convergence to a normal limit. Such a convergence typically does not hold almost surely, but at least in distribution.
(c) Urns that, even after a proper scaling, do not lead to convergence. Instead there is an asymptotic oscillatory behaviour of the distributions. Such oscillatory behaviour can even be captured almost surely; we discuss a (weak) description for distributions.

The cases of type (a) can be dealt with on the space $(\mathcal{M}_2^{\mathbb{R}}(\mu))^{\times d}$ with appropriate $\mu \in \mathbb{R}$ and $d \in \mathbb{N}$, where, by centering, one can always achieve the choice $\mu = 0$. One can use the metrics $\zeta_2^{\vee}$ or $\ell_2^{\vee}$, which lead to similar results although based on different details in the proofs. We will only present the use of $\zeta_2^{\vee}$, since we can then easily extend the argument to the type (b) cases by switching from $\zeta_2^{\vee}$ to $\zeta_3^{\vee}$. This leads to a more concise presentation. However, the $\ell_2^{\vee}$-metric appears to us to be equally convenient to apply in type (a) cases.

The cases of type (b) can be dealt with on the space $(\mathcal{M}_s^{\mathbb{R}}(\mu,\sigma^2))^{\times d}$ with $2 < s \leqslant 3$ and appropriate $\mu \in \mathbb{R}$, $\sigma > 0$ and $d \in \mathbb{N}$. By normalization, one can always achieve the choices $\mu = 0$ and $\sigma = 1$. Since in the context of urns third absolute moments in type (b) cases typically exist, one can use $s = 3$ and the metric $\zeta_3^{\vee}$. We do not know how to use the $\ell_p^{\vee}$-metrics in type (b) cases.

The cases of type (c) can be dealt with on the space $(\mathcal{M}_2^{\mathbb{C}}(\gamma))^{\times d}$ with appropriate $\gamma \in \mathbb{R}$ and $d \in \mathbb{N}$. The metric used subsequently in type (c) cases is the complex version of $\ell_2^{\vee}$. In our example below we will, however, use $\mathcal{M}_2^{\mathbb{C}}(\gamma_1) \times \cdots \times \mathcal{M}_2^{\mathbb{C}}(\gamma_d)$ with $\gamma_1,\dots,\gamma_d \in \mathbb{C}$ in order to be able to work with a more natural scaling of the random variables, the metric still being $\ell_2^{\vee}$. We think $\zeta_2^{\vee}$ can also be used in type (c) cases, but we have not checked the details since the application of $\ell_2^{\vee}$ is straightforward.

5. Associated fixed point equations

We fix $d, d' \in \mathbb{N}$, a $d \times d'$ matrix (A_{ir}) of random variables and a vector $(b_1,\dots,b_d)$ of random variables. Either all of these random variables are real or all of them are complex. Furthermore, we are given a $d \times d'$ matrix $(\pi(i,r))$ with all entries $\pi(i,r) \in \{1,\dots,d\}$. First,

https://doi.org/10.1017/S0963548314000364 Published online by Cambridge University Press

we consider the case where all A_{ir} and all b_i are real. We associate a map

$$T : (\mathcal{M}^{\mathbb{R}})^{\times d} \to (\mathcal{M}^{\mathbb{R}})^{\times d},$$
$$(\mu_1,\dots,\mu_d) \mapsto (T_1(\mu_1,\dots,\mu_d),\dots,T_d(\mu_1,\dots,\mu_d)), \tag{5.1}$$

$$T_i(\mu_1,\dots,\mu_d) := \mathcal{L}\left(\sum_{r=1}^{d'} A_{ir}Z_{ir} + b_i\right), \tag{5.2}$$

with $(A_{i1},\dots,A_{id'},b_i), Z_{i1},\dots,Z_{id'}$ independent and Z_{ir} distributed as $\mu_{\pi(i,r)}$, $r = 1,\dots,d'$ and for all components $i = 1,\dots,d$.

In the case where the A_{ir} and b_i are complex random variables, we define a map T' similar to T:

$$T' : (\mathcal{M}^{\mathbb{C}})^{\times d} \to (\mathcal{M}^{\mathbb{C}})^{\times d}, \tag{5.3}$$
$$(\mu_1,\dots,\mu_d) \mapsto (T'_1(\mu_1,\dots,\mu_d),\dots,T'_d(\mu_1,\dots,\mu_d)),$$

with $T'_i(\mu_1,\dots,\mu_d)$ defined as for T_i in (5.2).

For the three regimes discussed in the preview within Section 4 we use the following three theorems (Theorem 5.1 for type (a), Theorem 5.2 for type (b), and Theorem 5.3 for type (c)) on existence of fixed points of T and T'.

Theorem 5.1. *Assume that in the definition of T in (5.1) and (5.2), the A_{ir} and b_i are square-integrable real random variables with $\mathbb{E}[b_i] = 0$ for all $1 \leqslant i \leqslant d$ and $1 \leqslant r \leqslant d'$, and*

$$\max_{1\leqslant i\leqslant d} \sum_{r=1}^{d'} \mathbb{E}\left[A_{ir}^2\right] < 1. \tag{5.4}$$

Then the restriction of T to $(\mathcal{M}_2^{\mathbb{R}}(0))^{\times d}$ has a unique fixed point.

Theorem 5.2. *Assume that in the definition of T in (5.1) and (5.2) for some $\varepsilon > 0$, the A_{ir} are $L_{2+\varepsilon}$-integrable real random variables and $b_i = 0$ for all $1 \leqslant i \leqslant d$ and $1 \leqslant r \leqslant d'$, that almost surely*

$$\sum_{r=1}^{d'} A_{ir}^2 = 1 \quad \textit{for all } i = 1,\dots,d, \tag{5.5}$$

and

$$\min_{1\leqslant i\leqslant d} \mathbb{P}\left(\max_{1\leqslant r\leqslant d'} |A_{ir}| < 1\right) > 0. \tag{5.6}$$

Then, for all $\sigma^2 \geqslant 0$, the restriction of T to $(\mathcal{M}_{2+\varepsilon}^{\mathbb{R}}(0,\sigma^2))^{\times d}$ has the unique fixed point

$$(\mathcal{N}(0,\sigma^2),\dots,\mathcal{N}(0,\sigma^2)).$$

Theorem 5.3. *Assume that in the definition of T' in (5.3), the A_{ir} and b_i are square-integrable complex random variables for all $1 \leqslant i \leqslant d$ and $1 \leqslant r \leqslant d'$, and that for*

https://doi.org/10.1017/S0963548314000364 Published online by Cambridge University Press

$\gamma_1,\dots,\gamma_d \in \mathbb{C}$ we have

$$\mathbb{E}[b_i] + \sum_{r=1}^{d'} \gamma_{\pi(i,r)} \mathbb{E}[A_{ir}] = \gamma_i, \quad i = 1,\dots,d. \tag{5.7}$$

If, moreover,

$$\max_{1 \leqslant i \leqslant d} \sum_{r=1}^{d'} \mathbb{E}\big[|A_{ir}|^2\big] < 1, \tag{5.8}$$

then the restriction of T' to $\mathcal{M}_2^{\mathbb{C}}(\gamma_1) \times \cdots \times \mathcal{M}_2^{\mathbb{C}}(\gamma_d)$ has a unique fixed point.

Note that a special case of Theorem 5.1 was used in the proof of [16, Theorem 3.9(iii)] with a proof technique similar to that in our proof of Theorem 5.3.

The rest of this section contains the proofs of Theorems 5.1–5.3.

Proof of Theorem 5.1. First note that for $(\mu_1,\dots,\mu_d) \in (\mathcal{M}_2^{\mathbb{R}}(0))^{\times d}$, by independence in definition (5.2) and $\mathbb{E}[b_i] = 0$, we have $T_i(\mu_1,\dots,\mu_d) \in \mathcal{M}_2^{\mathbb{R}}(0)$ for $i = 1,\dots,d$. Hence, the restriction of T to $(\mathcal{M}_2^{\mathbb{R}}(0))^{\times d}$ maps into $(\mathcal{M}_2^{\mathbb{R}}(0))^{\times d}$.

Next, we show that the restriction of T to $(\mathcal{M}_2^{\mathbb{R}}(0))^{\times d}$ is a (strict) contraction with respect to the metric $\zeta_2^{\vee}$. For $(\mu_1,\dots,\mu_d),(\nu_1,\dots,\nu_d) \in (\mathcal{M}_2^{\mathbb{R}}(0))^{\times d}$ we first fix $i \in \{1,\dots,d\}$. Let $Z_{i1},\dots,Z_{id'}$ and $Z'_{i1},\dots,Z'_{id'}$ be real random variables such that Z_{ir} is distributed as $\mu_{\pi(i,r)}$ and Z'_{ir} is distributed as $\nu_{\pi(i,r)}$. Moreover, assume that both families

$$\{(A_{i1},\dots,A_{id'},b_i),\ Z_{i1},\dots,Z_{id'}\} \quad \text{and} \quad \{(A_{i1},\dots,A_{id'},b_i),\ Z'_{i1},\dots,Z'_{id'}\}$$

are independent. Then we have

$$T_i(\mu_1,\dots,\mu_d) = \mathcal{L}\left(\sum_{r=1}^{d'} A_{ir} Z_{ir} + b_i\right), \quad T_i(\nu_1,\dots,\nu_d) = \mathcal{L}\left(\sum_{r=1}^{d'} A_{ir} Z'_{ir} + b_i\right). \tag{5.9}$$

Conditioning on $(A_{i1},\dots,A_{id'},b_i)$ and denoting this vector's distribution by Υ, we obtain

$$\begin{aligned}
&\zeta_2(T_i(\mu_1,\dots,\mu_d), T_i(\nu_1,\dots,\nu_d)) \\
&\quad = \sup_{f \in \mathcal{F}_2} \left| \int \mathbb{E}\left[f\left(\sum_{r=1}^{d'} \alpha_r Z_{ir} + \beta\right) - f\left(\sum_{r=1}^{d'} \alpha_r Z'_{ir} + \beta\right)\right] d\Upsilon(\alpha_1,\dots,\alpha_{d'},\beta)\right| \\
&\quad \leqslant \int \sup_{f \in \mathcal{F}_2} \left| \mathbb{E}\left[f\left(\sum_{r=1}^{d'} \alpha_r Z_{ir} + \beta\right) - f\left(\sum_{r=1}^{d'} \alpha_r Z'_{ir} + \beta\right)\right]\right| d\Upsilon(\alpha_1,\dots,\alpha_{d'},\beta) \\
&\quad = \int \zeta_2\left(\sum_{r=1}^{d'} \alpha_r Z_{ir} + \beta, \sum_{r=1}^{d'} \alpha_r Z'_{ir} + \beta\right) d\Upsilon(\alpha_1,\dots,\alpha_{d'},\beta).
\end{aligned} \tag{5.10}$$

Since ζ_2 is $(2,+)$-ideal, we obtain from (4.4) that

$$\zeta_2\left(\sum \alpha_r Z_{ir} + \beta, \sum \alpha_r Z'_{ir} + \beta\right) \leqslant \sum \alpha_r^2 \zeta_2(Z_{ir}, Z'_{ir}).$$

Hence, we can further estimate

$$\begin{aligned}\zeta_2(T_i(\mu_1,\dots,\mu_d),\,T_i(\nu_1,\dots,\nu_d))&\\ &\leqslant \int \sum_{r=1}^{d'} \alpha_r^2 \zeta_2(Z_{ir}, Z'_{ir})\, d\Upsilon(\alpha_1,\dots,\alpha_{d'},\beta)\\ &= \int \sum_{r=1}^{d'} \alpha_r^2 \zeta_2(\mu_{\pi(i,r)}, \nu_{\pi(i,r)})\, d\Upsilon(\alpha_1,\dots,\alpha_{d'},\beta)\\ &\leqslant \left(\sum_{r=1}^{d'} \mathbb{E}\left[A_{ir}^2\right]\right) \zeta_2^{\vee}((\mu_1,\dots,\mu_d),(\nu_1,\dots,\nu_d)). \end{aligned} \tag{5.11}$$

Now, taking the maximum over i yields

$$\zeta_2^{\vee}(T(\mu_1,\dots,\mu_d), T(\nu_1,\dots,\nu_d)) \leqslant \left(\max_{1\leqslant i\leqslant d} \sum_{r=1}^{d'} \mathbb{E}\left[A_{ir}^2\right]\right) \zeta_2^{\vee}((\mu_1,\dots,\mu_d),(\nu_1,\dots,\nu_d)). \tag{5.12}$$

Hence, condition (5.4) implies that the restriction of T to $(\mathcal{M}_2^{\mathbb{R}}(0))^{\times d}$ is a contraction. Since the metric $\zeta_2^{\vee}$ is complete, Banach's fixed point theorem implies the assertion. □

Proof of Theorem 5.2. This proof is similar to the previous proof of Theorem 5.1. Let $\varepsilon > 0$ be as in Theorem 5.2 and let $\sigma > 0$ be arbitrary. First note that for

$$(\mu_1,\dots,\mu_d) \in (\mathcal{M}_{2+\varepsilon}^{\mathbb{R}}(0,\sigma^2))^{\times d},$$

by independence in definition (5.2), condition (5.5), and $b_i = 0$, we have

$$T_i(\mu_1,\dots,\mu_d) \in \mathcal{M}_{2+\varepsilon}^{\mathbb{R}}(0,\sigma^2) \quad \text{for } i = 1,\dots,d.$$

Hence, the restriction of T to $(\mathcal{M}_{2+\varepsilon}^{\mathbb{R}}(0,\sigma^2))^{\times d}$ maps into $(\mathcal{M}_{2+\varepsilon}^{\mathbb{R}}(0,\sigma^2))^{\times d}$.

We set $s := (2+\varepsilon) \wedge 3$. For

$$(\mu_1,\dots,\mu_d), (\nu_1,\dots,\nu_d) \in (\mathcal{M}_{2+\varepsilon}^{\mathbb{R}}(0,\sigma^2))^{\times d}$$

we choose $Z_{i1},\dots,Z_{id'}$ and $Z'_{i1},\dots,Z'_{id'}$ as in the proof of Theorem 5.1, such that we have (5.9). Note that with our choice of s we have

$$\zeta_s(T_i(\mu_1,\dots,\mu_d),\ T_i(\nu_1,\dots,\nu_d)) < \infty.$$

With an estimate analogous to (5.10)–(5.12), now using that ζ_s is $(s,+)$-ideal, we obtain

$$\zeta_s^{\vee}(T(\mu_1,\dots,\mu_d), T(\nu_1,\dots,\nu_d)) \leqslant \left(\max_{1\leqslant i\leqslant d} \sum_{r=1}^{d'} \mathbb{E}\left[|A_{ir}|^s\right]\right) \zeta_s^{\vee}((\mu_1,\dots,\mu_d),(\nu_1,\dots,\nu_d)).$$

Note that $s > 2$ and the conditions (5.5) and (5.6) imply that

$$\sum_{r=1}^{d'} \mathbb{E}[|A_{ir}|^s] < 1 \quad \text{for all } i = 1,\dots,d.$$

https://doi.org/10.1017/S0963548314000364 Published online by Cambridge University Press

Hence, the restriction of T to $(\mathcal{M}_{2+\varepsilon}^{\mathbb{R}}(0,\sigma^2))^{\times d}$ is a contraction and the completeness of $\zeta_s^\vee$ implies the existence of a unique fixed point. With the convolution property

$$\mathcal{N}(0,\sigma_1^2) * \mathcal{N}(0,\sigma_2^2) = \mathcal{N}(0,\sigma_1^2+\sigma_2^2) \quad \text{for } \sigma_1,\sigma_2 \geqslant 0,$$

one can directly check that $(\mathcal{N}(0,\sigma^2),\dots,\mathcal{N}(0,\sigma^2))$ is a fixed point of T in $(\mathcal{M}_{2+\varepsilon}^{\mathbb{R}}(0,\sigma^2))^{\times d}$. □

Proof of Theorem 5.3. Let $\gamma_1,\dots,\gamma_d$ be as in Theorem 5.3 and abbreviate

$$\mathcal{P} := \mathcal{M}_2^{\mathbb{C}}(\gamma_1) \times \cdots \times \mathcal{M}_2^{\mathbb{C}}(\gamma_d).$$

First note that for $(\mu_1,\dots,\mu_d) \in \mathcal{P}$, from independence in the definition of $T_i'(\mu_1,\dots,\mu_d)$ and the finite second moments of the A_{ir} and b_i, we obtain $T_i'(\mu_1,\dots,\mu_d) \in \mathcal{M}_2^{\mathbb{C}}$ for all $i = 1,\dots,d$. For a random variable W with distribution $T_i'(\mu_1,\dots,\mu_d)$, we have

$$\mathbb{E}[W] = \sum_{r=1}^{d'} \mathbb{E}[A_{ir}]\gamma_{\pi(i,r)} + \mathbb{E}[b_i] = \gamma_i$$

by condition (5.7). Hence, the restriction of T' to $\mathcal{P}$ maps into $\mathcal{P}$.

Next, we show that the restriction of T' to $\mathcal{P}$ is a contraction with respect to the metric $\ell_2^\vee$. For $(\mu_1,\dots,\mu_d),(\nu_1,\dots,\nu_d) \in \mathcal{P}$ we first fix $i \in \{1,\dots,d\}$. Let (Z_{ir},Z_{ir}') be an optimal coupling of $\mu_{\pi(i,r)}$ and $\nu_{\pi(i,r)}$ for $r = 1,\dots,d'$ such that $(Z_{i1},Z_{i1}'),\dots,(Z_{id'},Z_{id'}'),(A_{i1},\dots,A_{id'},b_i)$ are independent. Then we have

$$T_i'(\mu_1,\dots,\mu_d) = \mathcal{L}\left(\sum_{r=1}^{d'} A_{ir}Z_{ir} + b_i\right), \quad T_i'(\nu_1,\dots,\nu_d) = \mathcal{L}\left(\sum_{r=1}^{d'} A_{ir}Z_{ir}' + b_i\right). \tag{5.13}$$

Letting $\overline{\gamma}$ denote the complex conjugate of $\gamma \in \mathbb{C}$, we obtain

$$\begin{aligned}
&\ell_2^2(T_i'(\mu_1,\dots,\mu_d), T_i'(\nu_1,\dots,\nu_d)) \\
&\quad\leqslant \mathbb{E}\left[\left|\sum_{r=1}^{d'} A_{ir}(Z_{ir}-Z_{ir}')\right|^2\right] \\
&\quad= \mathbb{E}\left[\sum_{r=1}^{d'} |A_{ir}|^2|Z_{ir}-Z_{ir}'|^2\right] + \mathbb{E}\left[\sum_{r\neq t} A_{ir}(Z_{ir}-Z_{ir}')\overline{A_{it}(Z_{it}-Z_{it}')}\right] \\
&\quad= \sum_{r=1}^{d'} \mathbb{E}\left[|A_{ir}|^2\right]\ell_2^2(\mu_{\pi(i,r)},\nu_{\pi(i,r)}) \qquad (5.14) \\
&\quad\leqslant \left(\sum_{r=1}^{d'} \mathbb{E}\left[|A_{ir}|^2\right]\right)\left(\ell_2^\vee((\mu_1,\dots,\mu_d),(\nu_1,\dots,\nu_d))\right)^2.
\end{aligned}$$

For equation (5.14), we first use that $Z_{ir}-Z_{ir}'$ and $Z_{it}-Z_{it}'$ are independent, centred factors, so that the expectation of the sum over $r \neq t$ is 0, and second that (Z_{ir},Z_{ir}') are optimal couplings of $(\mu_{\pi(i,r)},\nu_{\pi(i,r)})$ such that $\mathbb{E}[|Z_{ir}-Z_{ir}'|^2] = \ell_2^2(\mu_{\pi(i,r)},\nu_{\pi(i,r)})$.

Now, taking the maximum over i yields

$$\ell_2^\vee(T'(\mu_1,\ldots,\mu_d),T'(\nu_1,\ldots,\nu_d)) \leqslant \left(\max_{1\leqslant i\leqslant d}\sum_{r=1}^{d'}\mathbb{E}\left[|A_{ir}|^2\right]\right)^{1/2}\ell_2^\vee((\mu_1,\ldots,\mu_d),(\nu_1,\ldots,\nu_d)).$$

Hence, condition (5.8) implies that the restriction of T' to $\mathcal{P}$ is a contraction. Since the metric $\ell_2^\vee$ is complete, Banach's fixed point theorem implies the assertion. □

6. Convergence and examples

In this section a couple of concrete Pólya urns are considered, and convergence of the normalized numbers of balls of a colour is shown within the product metrics defined in Section 4. The proofs are generic such that they can easily be transferred to other urns of types (a)–(c) in Section 3. We always show limit laws for the initial compositions of the urn with one ball of (arbitrary) colour. Limit laws for other initial compositions can be obtained from these by appropriate convolution with coefficients which are powers of components of an independent Dirichlet-distributed vector. We leave the details to the reader.

6.1. 2×2 deterministic replacement urns

A discussion of urns with a general balanced 2×2 replacement matrix as in (2.1) is given in Bagchi and Pal [3]. Subsequently, we assume the conditions in (2.1) and, as in [3], that $bc>0$. As shown in [3], asymptotic normal behaviour occurs for these urns when $a-c\leqslant (a+b)/2$ (type (b) in Section 4), whereas $a-c>(a+b)/2$ leads to limit laws with non-normal limit distributions (type (a) in Section 4). In this section we show how to derive these results by our contraction approach. With B_n^{b} and B_n^{w} as in the beginning of Section 2, we denote expectations by $\mu_{\mathrm{b}}(n)$ and $\mu_{\mathrm{w}}(n)$. These values can be derived exactly (see [3]):

$$\mu_{\mathrm{b}}(n)=\frac{c(a+b)}{b+c}n+\frac{b\,\Gamma\big(1/(a+b)\big)}{(b+c)\Gamma\big((1+a-c)/(a+b)\big)}\frac{\Gamma\big(n+(1+a-c)/(a+b)\big)}{\Gamma\big(n+1/(a+b)\big)}+\frac{c}{b+c}, \tag{6.1}$$

$$\mu_{\mathrm{w}}(n)=\frac{c(a+b)}{b+c}n-\frac{c\,\Gamma\big(1/(a+b)\big)}{(b+c)\Gamma\big((1+a-c)/(a+b)\big)}\frac{\Gamma\big(n+(1+a-c)/(a+b)\big)}{\Gamma\big(n+1/(a+b)\big)}+\frac{c}{b+c}. \tag{6.2}$$

Non-normal limit case. We first discuss the non-normal case $a-c>(a+b)/2$. Note that with $\lambda:=(a-c)/(a+b)$ and excluding the case $bc=0$, we have $1/2<\lambda<1$ and, as $n\to\infty$,

$$\mu_{\mathrm{b}}(n)=c_{\mathrm{b}}n+d_{\mathrm{b}}n^{\lambda}+o(n^{\lambda}),\quad \mu_{\mathrm{w}}(n)=c_{\mathrm{w}}n+d_{\mathrm{w}}n^{\lambda}+o(n^{\lambda}), \tag{6.3}$$

https://doi.org/10.1017/S0963548314000364 Published online by Cambridge University Press

with

$$\begin{aligned} c_{\mathrm{b}} &= c_{\mathrm{w}} = \frac{c(a+b)}{b+c}, \\ d_{\mathrm{b}} &= \frac{b\,\Gamma\big(1/(a+b)\big)}{(b+c)\Gamma\big((1+a-c)/(a+b)\big)}, \\ d_{\mathrm{w}} &= -\frac{c\,\Gamma\big(1/(a+b)\big)}{(b+c)\Gamma\big((1+a-c)/(a+b)\big)}. \end{aligned} \tag{6.4}$$

We use the normalizations $X_0 := Y_0 := 0$ and (see (3.1))

$$X_n := \frac{B_n^{\mathrm{b}} - \mu_{\mathrm{b}}(n)}{n^\lambda}, \quad Y_n := \frac{B_n^{\mathrm{w}} - \mu_{\mathrm{w}}(n)}{n^\lambda}, \quad n \geqslant 1. \tag{6.5}$$

Note that we do not have to identify the order of the variance in advance. It turns out that it is sufficient to use the order of the error terms $d_{\mathrm{b}} n^\lambda$ and $d_{\mathrm{w}} n^\lambda$ in the expansions (6.3). From the system (2.2)–(2.3) we obtain for the scaled quantities X_n, Y_n the following system for $n \geqslant 1$:

$$X_n \stackrel{d}{=} \sum_{r=1}^{a+1} \left(\frac{I_r^{(n)}}{n}\right)^\lambda X_{I_r^{(n)}}^{(r)} + \sum_{r=a+2}^{K} \left(\frac{I_r^{(n)}}{n}\right)^\lambda Y_{I_r^{(n)}}^{(r)} + b_{\mathrm{b}}(n), \tag{6.6}$$

$$Y_n \stackrel{d}{=} \sum_{r=1}^{c} \left(\frac{I_r^{(n)}}{n}\right)^\lambda X_{I_r^{(n)}}^{(r)} + \sum_{r=c+1}^{K} \left(\frac{I_r^{(n)}}{n}\right)^\lambda Y_{I_r^{(n)}}^{(r)} + b_{\mathrm{w}}(n), \tag{6.7}$$

with

$$b_{\mathrm{b}}(n) = d_{\mathrm{b}}\left(-1 + \sum_{r=1}^{a+1}\left(\frac{I_r^{(n)}}{n}\right)^\lambda\right) + d_{\mathrm{w}} \sum_{r=a+2}^{K}\left(\frac{I_r^{(n)}}{n}\right)^\lambda + o(1), \tag{6.8}$$

$$b_{\mathrm{w}}(n) = d_{\mathrm{b}} \sum_{r=1}^{c}\left(\frac{I_r^{(n)}}{n}\right)^\lambda + d_{\mathrm{w}}\left(-1 + \sum_{r=c+1}^{K}\left(\frac{I_r^{(n)}}{n}\right)^\lambda\right) + o(1), \tag{6.9}$$

with conditions on independence between the $X_j^{(r)}, Y_j^{(r)}$ and $I^{(n)}$ and identical distributions of the $X_j^{(r)}$ and $Y_j^{(r)}$ analogous to (2.2) and (2.3). The $o(1)$ terms in (6.8) and (6.9) are deterministic functions of $I^{(n)}$. In view of Lemma 2.1 this suggests, for limits X and Y of X_n and Y_n, respectively,

$$X \stackrel{d}{=} \sum_{r=1}^{a+1} D_r^\lambda X^{(r)} + \sum_{r=a+2}^{K} D_r^\lambda Y^{(r)} + b_{\mathrm{b}}, \tag{6.10}$$

$$Y \stackrel{d}{=} \sum_{r=1}^{c} D_r^\lambda X^{(r)} + \sum_{r=c+1}^{K} D_r^\lambda Y^{(r)} + b_{\mathrm{w}}, \tag{6.11}$$

with

$$b_{\rm b} = d_{\rm b}\left(-1 + \sum_{r=1}^{a+1} D_r^{\lambda}\right) + d_{\rm w} \sum_{r=a+2}^{K} D_r^{\lambda},$$

$$b_{\rm w} = d_{\rm b} \sum_{r=1}^{c} D_r^{\lambda} + d_{\rm w}\left(-1 + \sum_{r=c+1}^{K} D_r^{\lambda}\right),$$

where $(D_1,\dots,D_K)$, $X^{(1)},\dots,X^{(K)}$, $Y^{(1)},\dots,Y^{(K)}$ are independent, and the $X^{(r)}$ are distributed as X, the $Y^{(r)}$ are distributed as Y, and $(D_1,\dots,D_K)$ is as in Lemma 2.1. Note that the moments $\mathbb{E}[D_r^{\lambda}]$ and the form of $d_{\rm b}$ and $d_{\rm w}$ in (6.4) imply $\mathbb{E}[b_{\rm b}] = \mathbb{E}[b_{\rm w}] = 0$. From $\lambda > 1/2$ and $\sum_{r=1}^{K} D_r = 1$ we obtain

$$\sum_{r=1}^{K} \mathbb{E}\left[D_r^{2\lambda}\right] < 1.$$

Hence, Theorem 5.1 applies to the map associated to the system (6.10)–(6.11), and implies that there exists a unique solution $(\mathcal{L}(\Lambda_{\rm b}),\mathcal{L}(\Lambda_{\rm w}))$ in the space $\mathcal{M}_2^{\mathbb{R}}(0) \times \mathcal{M}_2^{\mathbb{R}}(0)$ to (6.10)–(6.11). The following convergence proof resembles ideas from Neininger and Rüschendorf [29].

Theorem 6.1. *Consider the Pólya urn with replacement matrix* (2.1) *with* $a-c > (a+b)/2$ *and* $bc > 0$*, and the normalized numbers* X_n *and* Y_n *of black balls as in* (6.5)*. Furthermore, let* $(\mathcal{L}(\Lambda_{\rm b}),\mathcal{L}(\Lambda_{\rm w}))$ *denote the unique solution of* (6.10)–(6.11) *in* $\mathcal{M}_2^{\mathbb{R}}(0) \times \mathcal{M}_2^{\mathbb{R}}(0)$*. Then, as* $n \to \infty$*,*

$$\zeta_2^{\vee}\big((X_n, Y_n),(\Lambda_{\rm b},\Lambda_{\rm w})\big) \to 0.$$

In particular, as $n \to \infty$*,*

$$X_n \xrightarrow{d} \Lambda_{\rm b}, \quad Y_n \xrightarrow{d} \Lambda_{\rm w}. \tag{6.12}$$

Proof. We first define, for $n \geqslant 1$, the accompanying sequences

$$Q_n^{\rm b} := \sum_{r=1}^{a+1} \left(\frac{I_r^{(n)}}{n}\right)^{\lambda} \Lambda_{\rm b}^{(r)} + \sum_{r=a+2}^{K} \left(\frac{I_r^{(n)}}{n}\right)^{\lambda} \Lambda_{\rm w}^{(r)} + b_{\rm b}(n), \tag{6.13}$$

$$Q_n^{\rm w} := \sum_{r=1}^{c} \left(\frac{I_r^{(n)}}{n}\right)^{\lambda} \Lambda_{\rm b}^{(r)} + \sum_{r=c+1}^{K} \left(\frac{I_r^{(n)}}{n}\right)^{\lambda} \Lambda_{\rm w}^{(r)} + b_{\rm w}(n), \tag{6.14}$$

with $b_{\rm b}(n)$ and $b_{\rm w}(n)$ as in (6.8) and the $\Lambda_{\rm b}^{(r)}$, $\Lambda_{\rm b}^{(r)}$ and $I^{(n)}$ being independent, where the $\Lambda_{\rm b}^{(r)}$ are distributed as $\Lambda_{\rm b}$ and the $\Lambda_{\rm w}^{(r)}$ are distributed as $\Lambda_{\rm w}$ for the respective values of r. Note that $Q_n^{\rm b}$ and $Q_n^{\rm w}$ are centred with finite second moments since $\mathcal{L}(\Lambda_{\rm b}),\mathcal{L}(\Lambda_{\rm b}) \in \mathcal{M}_2^{\mathbb{R}}(0)$. Hence, ζ_2 distances between $X_n, Y_n, Q_n^{\rm b}, Q_n^{\rm w}, \Lambda_{\rm b}$ and $\Lambda_{\rm w}$ are finite. To bound

$$\Delta(n) := \zeta_2^{\vee}((X_n, Y_n),(\Lambda_{\rm b},\Lambda_{\rm w})),$$

we look at the distances

$$\Delta_{\mathrm{b}}(n) := \zeta_2(X_n, \Lambda_{\mathrm{b}}), \quad \Delta_{\mathrm{w}}(n) := \zeta_2(Y_n, \Lambda_{\mathrm{w}}).$$

We start with the estimate

$$\zeta_2(X_n, \Lambda_{\mathrm{b}}) \leqslant \zeta_2(X_n, Q_n^{\mathrm{b}}) + \zeta_2(Q_n^{\mathrm{b}}, \Lambda_{\mathrm{b}}). \tag{6.15}$$

We first show for the second summand in the latter display that $\zeta_2(Q_n^{\mathrm{b}}, \Lambda_{\mathrm{b}}) \to 0$ as $n \to \infty$. With inequality (4.6), we have

$$\zeta_2(Q_n^{\mathrm{b}}, \Lambda_{\mathrm{b}}) \leqslant (\|Q_n^{\mathrm{b}}\|_2 + \|\Lambda_{\mathrm{b}}\|_2)\ell_2(Q_n^{\mathrm{b}}, \Lambda_{\mathrm{b}}).$$

Moreover, $\|\Lambda_{\mathrm{b}}\|_2 < \infty$ since $\mathcal{L}(\Lambda_{\mathrm{b}}) \in \mathcal{M}_2^{\mathbb{R}}$, and, by definition of Q_n^{b} and with $|I_r^{(n)}/n| \leqslant 1$, we have that $\|Q_n^{\mathrm{b}}\|_2$ is uniformly bounded in n. Hence, it is sufficient to show $\ell_2(Q_n^{\mathrm{b}}, \Lambda_{\mathrm{b}}) \to 0$. Using the independence properties in (6.13) and (6.10), we have that

$$\begin{aligned}&\ell_2(Q_n^{\mathrm{b}}, \Lambda_{\mathrm{b}})\\ &\quad\leqslant \sum_{r=1}^{a+1} \left\| \left(\frac{I_r^{(n)}}{n}\right)^{\lambda} - D_r^{\lambda} \right\|_2 \|\Lambda_{\mathrm{b}}^{(r)}\|_2 + \sum_{r=a+2}^{K} \left\| \left(\frac{I_r^{(n)}}{n}\right)^{\lambda} - D_r^{\lambda} \right\|_2 \|\Lambda_{\mathrm{w}}^{(r)}\|_2 + \|b_{\mathrm{b}}(n) - b_{\mathrm{b}}\|_2.\end{aligned}$$

Lemma 2.1 implies that

$$\|(I_r^{(n)}/n)^{\lambda} - D_r^{\lambda}\|_2 \to 0 \quad \text{as } n \to \infty,$$

which also implies $\|b_{\mathrm{b}}(n) - b_{\mathrm{b}}\|_2 \to 0$. Hence, we obtain

$$\ell_2(Q_n^{\mathrm{b}}, \Lambda_{\mathrm{b}}) \to 0 \quad \text{and} \quad \zeta_2(Q_n^{\mathrm{b}}, \Lambda_{\mathrm{b}}) \to 0.$$

Next, we bound the first summand $\zeta_2(X_n, Q_n^{\mathrm{b}})$ in (6.15). We condition on $I^{(n)}$. Note that conditionally on $I^{(n)}$ we have that $b_{\mathrm{b}}(n)$ is deterministic, which, for integration, we denote by $\beta = \beta(I^{(n)})$. Denoting the distribution of $I^{(n)}$ by Υ_n and $\mathbf{i} := (i_1, \dots, i_K)$, this yields

$$\begin{aligned}&\zeta_2(X_n, Q_n^{\mathrm{b}})\\ &\quad\leqslant \int \zeta_2\Bigg(\sum_{r=1}^{a+1}\left(\frac{i_r}{n}\right)^{\lambda} X_{i_r}^{(r)} + \sum_{r=a+2}^{K}\left(\frac{i_r}{n}\right)^{\lambda} Y_{i_r}^{(r)} + \beta,\\ &\qquad\qquad \sum_{r=1}^{a+1}\left(\frac{i_r}{n}\right)^{\lambda} \Lambda_{\mathrm{b}}^{(r)} + \sum_{r=a+2}^{K}\left(\frac{i_r}{n}\right)^{\lambda} \Lambda_{\mathrm{w}}^{(r)} + \beta\Bigg)\, d\Upsilon_n(\mathbf{i})\\ &\quad\leqslant \int \Bigg(\sum_{r=1}^{a+1}\left(\frac{i_r}{n}\right)^{2\lambda} \zeta_2(X_{i_r}^{(r)}, \Lambda_{\mathrm{b}}^{(r)}) + \sum_{r=a+2}^{K}\left(\frac{i_r}{n}\right)^{2\lambda} \zeta_2(Y_{i_r}^{(r)}, \Lambda_{\mathrm{w}}^{(r)})\Bigg)\, d\Upsilon_n(\mathbf{i})\\ &\quad= \sum_{r=1}^{a+1} \mathbb{E}\left[\left(\frac{I_r^{(n)}}{n}\right)^{2\lambda} \Delta_{\mathrm{b}}(I_r^{(n)})\right] + \sum_{r=a+2}^{K} \mathbb{E}\left[\left(\frac{I_r^{(n)}}{n}\right)^{2\lambda} \Delta_{\mathrm{w}}(I_r^{(n)})\right]\\ &\quad\leqslant \sum_{r=1}^{K} \mathbb{E}\left[\left(\frac{I_r^{(n)}}{n}\right)^{2\lambda} \Delta(I_r^{(n)})\right],\end{aligned} \tag{6.16}$$

https://doi.org/10.1017/S0963548314000364 Published online by Cambridge University Press

where for (6.16) we use that ζ_2 is $(2,+)$-ideal, as well as (4.5). Altogether, the estimate started in (6.15) yields

$$\Delta_{\mathrm{b}}(n) \leqslant \sum_{r=1}^{K} \mathbb{E}\left[\left(\frac{I_r^{(n)}}{n}\right)^{2\lambda} \Delta(I_r^{(n)})\right] + o(1).$$

With the same argument we obtain the same upper bound for $\Delta_{\mathrm{w}}(n)$. Thus, also using that $I_1^{(n)},\ldots,I_K^{(n)}$ are identically distributed, we have

$$\Delta(n) \leqslant K\mathbb{E}\left[\left(\frac{I_1^{(n)}}{n}\right)^{2\lambda} \Delta(I_1^{(n)})\right] + o(1). \tag{6.17}$$

Now, a standard argument implies $\Delta(n) \to 0$, as follows. First, from (6.17) we obtain with $I_1^{(n)}/n \to D_1$ in L_2 and, by $\lambda > 1/2$, with $\vartheta := K\mathbb{E}[D_1^{2\lambda}] < 1$ that

$$\begin{aligned}\Delta(n) &\leqslant K\mathbb{E}\left[\left(\frac{I_1^{(n)}}{n}\right)^{2\lambda}\right] \max_{0\leqslant k\leqslant n-1} \Delta(k) + o(1)\\ &\leqslant (\vartheta + o(1)) \max_{0\leqslant k\leqslant n-1} \Delta(k) + o(1).\end{aligned}$$

Since $\vartheta < 1$, this implies that the sequence $(\Delta(n))_{n\geqslant 0}$ is bounded. We denote $\eta := \sup_{n\geqslant 0} \Delta(n)$ and $\xi := \limsup_{n\to\infty} \Delta(n)$. For any $\varepsilon > 0$ there exists an $n_0 \geqslant 0$ such that $\Delta(n) \leqslant \xi + \varepsilon$ for all $n \geqslant n_0$. Hence, from (6.17) we obtain

$$\Delta(n) \leqslant K\mathbb{E}\left[\mathbf{1}_{\{I_1^{(n)}<n_0\}}\left(\frac{I_1^{(n)}}{n}\right)^{2\lambda}\right]\eta + K\mathbb{E}\left[\mathbf{1}_{\{I_1^{(n)}\geqslant n_0\}}\left(\frac{I_1^{(n)}}{n}\right)^{2\lambda}\right](\xi+\varepsilon) + o(1).$$

With $n \to \infty$ this implies

$$\xi \leqslant \vartheta(\xi + \varepsilon).$$

Since $\vartheta < 1$ and $\varepsilon > 0$ is arbitrary, this implies $\xi = 0$. Hence, we have

$$\zeta_2^{\vee}((X_n, Y_n), (\Lambda_{\mathrm{b}}, \Lambda_{\mathrm{w}})) \to 0 \quad \text{as } n \to \infty.$$

Since convergence in ζ_2 implies weak convergence, this implies (6.12) too. □

The normal limit case. Now we discuss the normal limit case $a - c \leqslant (a+b)/2$, where we first consider $a - c < (a+b)/2$. (The remaining case $a - c = (a+b)/2$ is similar with more involved expansions for the first two moments.) The formulae (6.1), (6.2) now imply

$$\mu_{\mathrm{b}}(n) = c_{\mathrm{b}} n + o(\sqrt{n}), \quad \mu_{\mathrm{w}}(n) = c_{\mathrm{w}} n + o(\sqrt{n}), \tag{6.18}$$

with c_{b} and c_{w} as in (6.4). As usual in the use of the contraction method for proving normal limit laws based on the metric ζ_3, we also need an expansion of the variance. We denote the variances of B_n^{b} and B_n^{w} by $\sigma_{\mathrm{b}}^2(n)$ and $\sigma_{\mathrm{w}}^2(n)$. As well as $bc = 0$, we exclude the case $a = c$. (In this case there is a trivial non-random evolution of the urn.) From [3] we have as $n \to \infty$:

$$\sigma_{\mathrm{b}}^2(n) = f_{\mathrm{b}} n + o(n), \quad \sigma_{\mathrm{w}}^2(n) = f_{\mathrm{w}} n + o(n), \tag{6.19}$$

https://doi.org/10.1017/S0963548314000364 Published online by Cambridge University Press

with

$$f_{\mathrm{b}} = f_{\mathrm{w}} = \frac{(a+b)bc(a-c)^2}{(a+b-2(a-c))(b+c)^2} > 0.$$

We use the normalizations $X_0 := Y_0 := X_1 := Y_1 := 0$ and (see (3.3))

$$X_n := \frac{B_n^{\mathrm{b}} - \mu_{\mathrm{b}}(n)}{\sigma_{\mathrm{b}}(n)}, \quad Y_n := \frac{B_n^{\mathrm{w}} - \mu_{\mathrm{w}}(n)}{\sigma_{\mathrm{w}}(n)}, \quad n \geqslant 2. \tag{6.20}$$

From the system (2.2)–(2.3) we obtain for the scaled quantities X_n, Y_n, for $n \geqslant 1$, the system

$$X_n \stackrel{d}{=} \sum_{r=1}^{a+1} \frac{\sigma_{\mathrm{b}}(I_r^{(n)})}{\sigma_{\mathrm{b}}(n)} X_{I_r^{(n)}}^{(r)} + \sum_{r=a+2}^{K} \frac{\sigma_{\mathrm{w}}(I_r^{(n)})}{\sigma_{\mathrm{b}}(n)} Y_{I_r^{(n)}}^{(r)} + e_{\mathrm{b}}(n), \tag{6.21}$$

$$Y_n \stackrel{d}{=} \sum_{r=1}^{c} \frac{\sigma_{\mathrm{b}}(I_r^{(n)})}{\sigma_{\mathrm{w}}(n)} X_{I_r^{(n)}}^{(r)} + \sum_{r=c+1}^{K} \frac{\sigma_{\mathrm{w}}(I_r^{(n)})}{\sigma_{\mathrm{w}}(n)} Y_{I_r^{(n)}}^{(r)} + e_{\mathrm{w}}(n), \tag{6.22}$$

with conditions on independence and identical distributions analogous to (2.2) and (2.3) (respectively (6.6) and (6.7)). We have $\|e_{\mathrm{b}}(n)\|_\infty, \|e_{\mathrm{w}}(n)\|_\infty \to 0$ since the leading linear terms in the expansions (6.18) cancel out and the error terms $o(\sqrt{n})$ are asymptotically eliminated by the scaling of order $1/\sqrt{n}$. In view of Lemma 2.1, this suggests, for limits X and Y of X_n and Y_n, respectively,

$$X \stackrel{d}{=} \sum_{r=1}^{a+1} \sqrt{D_r} X^{(r)} + \sum_{r=a+2}^{K} \sqrt{D_r} Y^{(r)}, \tag{6.23}$$

$$Y \stackrel{d}{=} \sum_{r=1}^{c} \sqrt{D_r} X^{(r)} + \sum_{r=c+1}^{K} \sqrt{D_r} Y^{(r)}, \tag{6.24}$$

where $(D_1, \ldots, D_K)$, $X^{(1)}, \ldots, X^{(K)}$, $Y^{(1)}, \ldots, Y^{(K)}$ are independent, and the $X^{(r)}$ are distributed as X and the $Y^{(r)}$ are distributed as Y. We can apply Theorem 5.2 to the map associated to the system (6.23)–(6.24). The conditions (5.5) and (5.6) are trivially satisfied. Hence $(\mathcal{N}(0,1), \mathcal{N}(0,1))$ is the unique fixed point of the associated map in the space $\mathcal{M}_3^{\mathbb{R}}(0,1) \times \mathcal{M}_3^{\mathbb{R}}(0,1)$.

Theorem 6.2. *Consider the Pólya urn with replacement matrix* (2.1) *with* $a - c < (a+b)/2$ *and* $bc > 0$ *and the normalized numbers* X_n *and* Y_n *of black balls as in* (6.20). *Then, as* $n \to \infty$,

$$\zeta_3^{\vee}\big((X_n, Y_n), (\mathcal{N}(0,1), \mathcal{N}(0,1))\big) \to 0.$$

In particular, as $n \to \infty$,

$$X_n \stackrel{d}{\longrightarrow} \mathcal{N}(0,1), \quad Y_n \stackrel{d}{\longrightarrow} \mathcal{N}(0,1).$$

Proof. The proof of this theorem can be follow the approach of the proof of Theorem 6.1. However, more care has to be taken in the definition of the quantities corresponding to

https://doi.org/10.1017/S0963548314000364 Published online by Cambridge University Press

Q_n^{b} and Q_n^{w} in (6.13) in order to ensure finiteness of the ζ_3 distances. For $n \geqslant 2$, a possible choice is

$$\tilde{Q}_n^{\mathrm{b}} := \sum_{r=1}^{a+1} \mathbf{1}_{\{I_r^{(n)} \geqslant 2\}} \frac{\sigma_{\mathrm{b}}(I_r^{(n)})}{\sigma_{\mathrm{b}}(n)} N_r + \sum_{r=a+2}^{K} \mathbf{1}_{\{I_r^{(n)} \geqslant 2\}} \frac{\sigma_{\mathrm{w}}(I_r^{(n)})}{\sigma_{\mathrm{b}}(n)} N_r + e_{\mathrm{b}}(n), \tag{6.25}$$

$$\tilde{Q}_n^{\mathrm{w}} \stackrel{d}{=} \sum_{r=1}^{c} \mathbf{1}_{\{I_r^{(n)} \geqslant 2\}} \frac{\sigma_{\mathrm{b}}(I_r^{(n)})}{\sigma_{\mathrm{w}}(n)} N_r + \sum_{r=c+1}^{K} \mathbf{1}_{\{I_r^{(n)} \geqslant 2\}} \frac{\sigma_{\mathrm{w}}(I_r^{(n)})}{\sigma_{\mathrm{w}}(n)} N_r + e_{\mathrm{w}}(n), \tag{6.26}$$

with $e_{\mathrm{b}}(n)$ and $e_{\mathrm{w}}(n)$ as in (6.21)–(6.22) and $N_1, \dots, N_K$, $I^{(n)}$, independent, where the N_r are standard normally distributed for $r = 1, \dots, K$. A comparison of the definition of $\tilde{Q}_n^{\mathrm{b}}$ and $\tilde{Q}_n^{\mathrm{w}}$ with the right-hand sides of (6.21) and (6.22) and the scaling (6.20) yields $\mathbb{E}[\tilde{Q}_n^{\mathrm{b}}] = \mathbb{E}[\tilde{Q}_n^{\mathrm{w}}] = 0$ and $\mathrm{Var}(\tilde{Q}_n^{\mathrm{b}}) = \mathrm{Var}(\tilde{Q}_n^{\mathrm{w}}) = 1$ for all $n \geqslant 2$. Obviously, we also have $\|\tilde{Q}_n^{\mathrm{b}}\|_3, \|\tilde{Q}_n^{\mathrm{w}}\|_3 < \infty$. Hence, ζ_3 distances between X_n, Y_n, $\tilde{Q}_n^{\mathrm{b}}$, $\tilde{Q}_n^{\mathrm{w}}$, and $\mathcal{N}(0,1)$ are finite for all $n \geqslant 2$. With

$$\begin{aligned}
\tilde{\Delta}(n) &:= \zeta_3^{\vee}((X_n, Y_n), (\mathcal{N}(0,1), \mathcal{N}(0,1))),\\
\tilde{\Delta}_{\mathrm{b}}(n) &:= \zeta_3(X_n, \mathcal{N}(0,1)),\\
\tilde{\Delta}_{\mathrm{w}}(n) &:= \zeta_3(Y_n, \mathcal{N}(0,1)),
\end{aligned}$$

we also start with

$$\zeta_3(X_n, \mathcal{N}(0,1)) \leqslant \zeta_3(X_n, \tilde{Q}_n^{\mathrm{b}}) + \zeta_3(\tilde{Q}_n^{\mathrm{b}}, \mathcal{N}(0,1)).$$

Analogous to the proof of Theorem 6.1, we obtain $\zeta_3(\tilde{Q}_n^{\mathrm{b}}, \mathcal{N}(0,1)) \to 0$ as $n \to \infty$.

The bound for $\zeta_3(X_n, \tilde{Q}_n^{\mathrm{b}})$ is also analogous to the proof of Theorem 6.1, where we use that ζ_3 is $(3,+)$-ideal instead of $(2,+)$-ideal. This yields

$$\zeta_3(X_n, \tilde{Q}_n^{\mathrm{b}}) \leqslant \sum_{r=1}^{a+1} \mathbb{E}\left[\left(\frac{\sigma_{\mathrm{b}}(I_r^{(n)})}{\sigma_{\mathrm{b}}(n)}\right)^3 \tilde{\Delta}(I_r^{(n)})\right] + \sum_{r=a+2}^{K} \mathbb{E}\left[\left(\frac{\sigma_{\mathrm{w}}(I_r^{(n)})}{\sigma_{\mathrm{b}}(n)}\right)^3 \tilde{\Delta}(I_r^{(n)})\right].$$

Then we argue as in the previous proof to obtain, analogous to (6.17),

$$\tilde{\Delta}(n) \leqslant \sum_{r=1}^{a+1} \mathbb{E}\left[\left(\frac{\sigma_{\mathrm{b}}(I_r^{(n)})}{\sigma_{\mathrm{b}}(n)}\right)^3 \tilde{\Delta}\big(I_r^{(n)}\big)\right] + \sum_{r=a+2}^{K} \mathbb{E}\left[\left(\frac{\sigma_{\mathrm{w}}(I_r^{(n)})}{\sigma_{\mathrm{b}}(n)}\right)^3 \tilde{\Delta}\big(I_r^{(n)}\big)\right] + o(1).$$

From this estimate we can deduce $\tilde{\Delta}(n) \to 0$ as for $\Delta(n)$ in the proof of Theorem 6.1, where we need to use the fact that from the expansions (6.19) and Lemma 2.1 we obtain, as $n \to \infty$, that

$$\sum_{r=1}^{a+1} \mathbb{E}\left[\left(\frac{\sigma_{\mathrm{b}}\big(I_r^{(n)}\big)}{\sigma_{\mathrm{b}}(n)}\right)^3\right] + \sum_{r=a+2}^{K} \mathbb{E}\left[\left(\frac{\sigma_{\mathrm{w}}\big(I_r^{(n)}\big)}{\sigma_{\mathrm{b}}(n)}\right)^3\right] \to \sum_{r=1}^{K} \mathbb{E}\big[D_r^{3/2}\big] < 1. \tag{6.27}$$

□

https://doi.org/10.1017/S0963548314000364 Published online by Cambridge University Press

Remarks. **(1)** Note that the proof of Theorem 6.2 is not suitable for the $\zeta_2^\vee$-metric since the term corresponding to (6.27) is then

$$\sum_{r=1}^{a+1}\mathbb{E}\left[\left(\frac{\sigma_{\mathrm{b}}(I_r^{(n)})}{\sigma_{\mathrm{b}}(n)}\right)^2\right]+\sum_{r=a+2}^{K}\mathbb{E}\left[\left(\frac{\sigma_{\mathrm{w}}(I_r^{(n)})}{\sigma_{\mathrm{b}}(n)}\right)^2\right]\to\sum_{r=1}^{K}\mathbb{E}\left[D_r\right]=1,$$

where a limit <1 is required to obtain $\tilde{\Delta}(n)\to 0$. This is why we use $\zeta_3^\vee$. It is possible to use $\zeta_s^\vee$ for any $2<s\leqslant 3$ leading to the limit $\sum_{r=1}^{K}\mathbb{E}[D_r^s]<1$.

(2) The case $a-c=(a+b)/2$ differs in the error terms in (6.18), which then become $\mathrm{O}(\sqrt{n})$. Since the variances in (6.19) get additional logarithmic factors, we still obtain the system (6.23)–(6.24), and our proof technique can be applied as well.

(3) The condition $bc>0$ cannot be dropped. In the case $bc=0$, the urn model is not irreducible in the terminology of Janson [16] and is known to behave quite differently. A comprehensive study of the case $bc=0$ is given in Janson [17]; see also Janson [19]. In our approach $bc=0$ would lead to degenerate systems of limit equations that do not identify limit laws.

(4) The condition $f_b=f_w$ is necessary for our proof to work.

6.2. An urn with random replacements

As an example of random entries in the replacement matrix R, we consider a simple model with two colours, black and white. In each step when a black ball is drawn, a coin is independently tossed to decide whether the black ball is placed back together with another black ball or together with another white ball. The probability of success (a second black ball) is denoted by $0<\alpha<1$. Similarly, if a white ball is drawn, a coin with probability $0<\beta<1$ is tossed to decide whether a second white ball or a black ball is placed back together with the white ball. We denote the replacement matrix by

$$R=\begin{bmatrix}F_\alpha & 1-F_\alpha\\ 1-F_\beta & F_\beta\end{bmatrix}, \tag{6.28}$$

where F_α and F_β denote Bernoulli random variables being 1 with probabilities α and β respectively, otherwise 0. This urn model was introduced in the context of clinical trials and studied together with generalizations in [37, 38, 36, 35, 27, 4, 5, 16].

The row sums of R in (6.28) are both almost surely equal to one, hence the urn is balanced. Again, the number of black balls after n draws starting with an initial composition with one black ball is denoted by B_n^{b}, and if starting with a white ball by B_n^{w}. According to our approach in Section 2 we obtain the recursive equation

$$B_n^{\mathrm{b}}\stackrel{d}{=}B_{I_n}^{\mathrm{b},(1)}+F_\alpha B_{J_n}^{\mathrm{b},(2)}+(1-F_\alpha)B_{J_n}^{\mathrm{w}},\quad n\geqslant 1, \tag{6.29}$$

where $(B_k^{\mathrm{b},(1)})_{0\leqslant k<n}$, $(B_k^{\mathrm{b},(2)})_{0\leqslant k<n}$, $(B_k^{\mathrm{w}})_{0\leqslant k<n}$, F_α and I_n are independent, and $B_k^{\mathrm{b},(1)}$ and $B_k^{\mathrm{b},(2)}$ are distributed as B_k^{b} for $k=0,\dots,n-1$, and I_n is uniformly distributed on $\{0,\dots,n-1\}$ while $J_n:=n-1-I_n$. (The uniform distribution of I_n follows from the uniform distribution of the number of balls in the $\left[\begin{smallmatrix}1&0\\0&1\end{smallmatrix}\right]$-Pólya urn.) Similarly, we obtain for B_n^{w}

https://doi.org/10.1017/S0963548314000364 Published online by Cambridge University Press

that

$$B_n^{\mathrm{w}} \stackrel{d}{=} B_{I_n}^{\mathrm{w},(1)} + F_\beta B_{J_n}^{\mathrm{w},(2)} + (1-F_\beta) B_{J_n}^{\mathrm{b}}, \quad n \geqslant 1, \tag{6.30}$$

with conditions on independence and identical distributions similar to (6.29). Together with the initial value $(B_0^{\mathrm{b}}, B_0^{\mathrm{w}}) = (1,0)$, the system of equations (6.29)–(6.30) again defines the sequence of pairs of distributions $(\mathcal{L}(B_n^{\mathrm{b}}), \mathcal{L}(B_n^{\mathrm{w}}))_{n\geqslant 0}$. As a special case of Lemma 2.1 we have

$$\left(\frac{I_n}{n}, \frac{J_n}{n}\right) \to (U, 1-U), \quad (n\to\infty), \tag{6.31}$$

almost surely where U is uniformly distributed on $[0,1]$. Furthermore, we denote for $n \geqslant 0$

$$\mu_{\mathrm{b}}(n) := \mathbb{E}[B_n^{\mathrm{b}}], \quad \mu_{\mathrm{w}}(n) := \mathbb{E}[B_n^{\mathrm{w}}]. \tag{6.32}$$

These means have been studied before. We have the following exact formulae.

Lemma 6.3. *For $\mu_{\mathrm{b}}(n)$ and $\mu_{\mathrm{w}}(n)$ as in (6.32) with $0<\alpha,\beta<1$, we have*

$$\mu_{\mathrm{b}}(n) = \frac{1-\beta}{2-\alpha-\beta} n + \frac{1-\alpha}{2-\alpha-\beta} \frac{\Gamma(n+\alpha+\beta)}{\Gamma(\alpha+\beta)\Gamma(n+1)} + \frac{1-\beta}{2-\alpha-\beta}, \tag{6.33}$$

$$\mu_{\mathrm{w}}(n) = \frac{1-\beta}{2-\alpha-\beta} n - \frac{1-\beta}{2-\alpha-\beta} \frac{\Gamma(n+\alpha+\beta)}{\Gamma(\alpha+\beta)\Gamma(n+1)} + \frac{1-\beta}{2-\alpha-\beta}. \tag{6.34}$$

Proof. The proof is based on matrix diagonalization and can easily be done along the lines of the proof of Lemma 6.7 below. □

As in the example from Section 6.1, we have two different types of limit laws, with normal limit for $\alpha+\beta \leqslant 3/2$ and non-normal limit for $\alpha+\beta > 3/2$.

The non-normal limit case. We assume that $\lambda := \alpha+\beta-1 > 1/2$. From Lemma 6.3 we obtain the asymptotic expressions, as $n\to\infty$,

$$\mu_{\mathrm{b}}(n) = c_{\mathrm{b}}' n + d_{\mathrm{b}}' n^\lambda + o(n^\lambda),$$
$$\mu_{\mathrm{w}}(n) = c_{\mathrm{w}}' n + d_{\mathrm{w}}' n^\lambda + o(n^\lambda),$$

with constants

$$c_{\mathrm{b}}' = c_{\mathrm{w}}' = \frac{1-\beta}{1-\lambda}, \quad d_{\mathrm{b}}' = \frac{1-\alpha}{(1-\lambda)\Gamma(\lambda+1)}, \quad d_{\mathrm{w}}' = -\frac{1-\beta}{(1-\lambda)\Gamma(\lambda+1)}. \tag{6.35}$$

We use the normalizations $X_0 := Y_0 := 0$ and (see (3.1))

$$X_n := \frac{B_n^{\mathrm{b}} - \mu_{\mathrm{b}}(n)}{n^\lambda}, \quad Y_n := \frac{B_n^{\mathrm{w}} - \mu_{\mathrm{w}}(n)}{n^\lambda}, \quad n \geqslant 1. \tag{6.36}$$

As in the non-normal case of the example in Section 6.1, it is sufficient to use the order of the error term of the mean for the scaling. From (6.29)–(6.30), we obtain for $n \geqslant 1$

$$X_n \stackrel{d}{=} \left(\frac{I_n}{n}\right)^\lambda X_{I_n}^{(1)} + F_\alpha \left(\frac{J_n}{n}\right)^\lambda X_{J_n}^{(2)} + (1-F_\alpha)\left(\frac{J_n}{n}\right)^\lambda Y_{J_n} + b_{\mathrm{b}}'(n), \tag{6.37}$$

$$Y_n \stackrel{d}{=} \left(\frac{I_n}{n}\right)^\lambda Y_{I_n}^{(1)} + F_\beta \left(\frac{J_n}{n}\right)^\lambda Y_{J_n}^{(2)} + (1-F_\beta)\left(\frac{J_n}{n}\right)^\lambda X_{J_n} + b_{\mathrm{w}}'(n), \tag{6.38}$$

with

$$b'_{\rm b}(n) = d'_{\rm b}\left(\left(\frac{I_n}{n}\right)^{\lambda} + F_\alpha\left(\frac{J_n}{n}\right)^{\lambda} - 1\right) + d'_{\rm w}(1-F_\alpha)\left(\frac{J_n}{n}\right)^{\lambda} + o(1),$$
$$b'_{\rm w}(n) = d'_{\rm w}\left(\left(\frac{I_n}{n}\right)^{\lambda} + F_\beta\left(\frac{J_n}{n}\right)^{\lambda} - 1\right) + d'_{\rm b}(1-F_\beta)\left(\frac{J_n}{n}\right)^{\lambda} + o(1),$$

with conditions on independence and identical distributions analogously to (6.29)–(6.30). In view of (6.31), this suggests, for limits X and Y of X_n and Y_n, that

$$X \stackrel{d}{=} U^\lambda X^{(1)} + F_\alpha(1-U)^\lambda X^{(2)} + (1-F_\alpha)(1-U)^\lambda Y^{(1)} + b'_{\rm b}, \tag{6.39}$$
$$Y \stackrel{d}{=} U^\lambda Y^{(1)} + F_\beta(1-U)^\lambda Y^{(2)} + (1-F_\beta)(1-U)^\lambda X^{(1)} + b'_{\rm w}, \tag{6.40}$$

with

$$b'_{\rm b} = d'_{\rm b}\big(U^\lambda + F_\alpha(1-U)^\lambda - 1\big) + d'_{\rm w}(1-F_\alpha)(1-U)^\lambda,$$
$$b'_{\rm w} = d'_{\rm w}\big(U^\lambda + F_\beta(1-U)^\lambda - 1\big) + d'_{\rm b}(1-F_\beta)(1-U)^\lambda,$$

where $X^{(1)}$, $X^{(2)}$, $Y^{(1)}$, $Y^{(2)}$ and U are independent and $X^{(1)}$, $X^{(2)}$ are distributed as X and $Y^{(1)}$, $Y^{(2)}$ are distributed as Y.

To check that Theorem 5.1 can be applied to the map associated to the system (6.39)–(6.40), first note that the form of $d'_{\rm b}$ and $d'_{\rm w}$ in (6.35) implies $\mathbb{E}[b'_{\rm b}] = \mathbb{E}[b'_{\rm w}] = 0$. To check condition (5.4), note that we have

$$\mathbb{E}\big[U^{2\lambda}\big] + \mathbb{E}\big[F_\alpha(1-U)^{2\lambda}\big] + \mathbb{E}\big[(1-F_\alpha)(1-U)^{2\lambda}\big] = \frac{2}{2\lambda+1} < 1,$$

since $\lambda > 1/2$. Analogously, we have $\mathbb{E}[U^{2\lambda}] + \mathbb{E}[F_\beta(1-U)^{2\lambda}] + \mathbb{E}[(1-F_\beta)(1-U)^{2\lambda}] = 2/(2\lambda+1) < 1$. Together, this verifies condition (5.4). Hence Theorem 5.1 can be applied, and yields a unique fixed point $(\mathcal{L}(\Lambda'_{\rm b}), \mathcal{L}(\Lambda'_{\rm w}))$ in $\mathcal{M}_2^{\mathbb{R}}(0) \times \mathcal{M}_2^{\mathbb{R}}(0)$ to (6.39)–(6.40).

Theorem 6.4. *Consider the Pólya urn with random replacement matrix* (6.28) *with* $\alpha, \beta \in (0,1)$ *and* $\alpha+\beta > 3/2$ *and the normalized numbers* X_n *and* Y_n *of black balls as in* (6.36). *Furthermore, let* $(\mathcal{L}(\Lambda'_{\rm b}), \mathcal{L}(\Lambda'_{\rm w}))$ *denote the unique solution of* (6.39)–(6.40) *in* $\mathcal{M}_2^{\mathbb{R}}(0) \times \mathcal{M}_2^{\mathbb{R}}(0)$. *Then, as* $n \to \infty$,

$$X_n \stackrel{d}{\longrightarrow} \Lambda'_{\rm b}, \quad Y_n \stackrel{d}{\longrightarrow} \Lambda'_{\rm w}.$$

Proof. The proof is analogous to that of Theorem 6.1. □

The normal limit case. Now we discuss the normal limit case $\lambda := \alpha+\beta-1 \leqslant 1/2$. We first assume $\lambda := \alpha+\beta-1 < 1/2$. The expansions from Lemma 6.3 now imply, as $n \to \infty$,

$$\mu_{\rm b}(n) = c_{\rm b} n + o(\sqrt{n}), \quad \mu_{\rm w}(n) = c_{\rm w} n + o(\sqrt{n}), \tag{6.41}$$

with $c_{\rm b}$ and $c_{\rm w}$ given in (6.35). As in the normal limit cases in the examples in Section 6.1, we first need asymptotic expressions for the variances. We denote the variances of $B_n^{\rm b}$ and $B_n^{\rm w}$ by $\hat\sigma_{\rm b}^2(n)$ and $\hat\sigma_{\rm w}^2(n)$. These can be obtained from a result of Matthews and Rosenberger [27] for the number of draws of each colour, as follows.

https://doi.org/10.1017/S0963548314000364 Published online by Cambridge University Press

Lemma 6.5. *We have, as $n \to \infty$,*

$$\hat{\sigma}_{\mathrm{b}}^2(n) = f'_{\mathrm{b}} n + o(n), \quad \hat{\sigma}_{\mathrm{w}}^2(n) = f'_{\mathrm{w}} n + o(n), \tag{6.42}$$

with

$$f'_{\mathrm{b}} = f'_{\mathrm{w}} = \frac{(1-\alpha)(1-\beta)}{(1-\lambda)^2}\left(\frac{1}{1-2\lambda} - 2\lambda(1+\lambda)\right) > 0.$$

Proof. Matthews and Rosenberger [27], for the present urn model, study the number N_n of draws within the first n draws in which a black ball is drawn. Starting with one black ball, they establish, as $n \to \infty$, that

$$\begin{aligned}\mathbb{E}\left[N_n\right] &= \frac{1-\beta}{1-\lambda}\, n + o(n),\\ \mathrm{Var}(N_n), &= \frac{(1-\alpha)(1-\beta)(3+2\lambda)}{(1-\lambda)^2(1-2\lambda)}\, n + o(n).\end{aligned}$$

As each black ball in the urn is either the first ball or has been added after drawing a black ball and having success in tossing the corresponding coin, or after drawing a white ball and having no success in tossing the coin, we can directly link N_n to B_n^{b}. Letting $(F_j^{\mathrm{b}})_{1 \leqslant j \leqslant N_n}$ denote the coin flips after drawing black balls and $(F_j^{\mathrm{w}})_{1 \leqslant j \leqslant (n-N_n)}$ denote the coin flips after drawing white balls, we have

$$B_n^{\mathrm{b}} = 1 + \sum_{j=1}^{N_n} F_j^{\mathrm{b}} + \sum_{j=1}^{n-N_n} (1 - F_j^{\mathrm{w}}).$$

Using that all coin flips are independent, we obtain from the law of total variance by conditioning on N_n that

$$\begin{aligned}\hat{\sigma}_b^2(n) &= \mathbb{E}\left[\mathrm{Var}\left(B_n^{\mathrm{b}} \mid N_n\right)\right] + \mathrm{Var}\left(\mathbb{E}\left[B_n^{\mathrm{b}} \mid N_n\right]\right)\\ &= \frac{(1-\alpha)(1-\beta)}{(1-\lambda)^2}\left(\frac{1}{1-2\lambda} - 2\lambda(1+\lambda)\right) n + o(n).\end{aligned}$$

When starting with one white ball, a similar argument gives the corresponding result. □

We use the normalizations $X_0 := Y_0 := 0$ and (see (3.3))

$$X_n := \frac{B_n^{\mathrm{b}} - \mu_{\mathrm{b}}(n)}{\hat{\sigma}_{\mathrm{b}}(n)}, \quad Y_n := \frac{B_n^{\mathrm{w}} - \mu_{\mathrm{w}}(n)}{\hat{\sigma}_{\mathrm{w}}(n)}, \quad n \geqslant 1. \tag{6.43}$$

From the system (6.29)–(6.30) we obtain for the scaled quantities X_n, Y_n, for $n \geqslant 1$, the system

$$\begin{aligned}X_n &\stackrel{d}{=} \frac{\hat{\sigma}_{\mathrm{b}}(I_n)}{\hat{\sigma}_{\mathrm{b}}(n)} X_{I_n}^{(1)} + F_\alpha \frac{\hat{\sigma}_{\mathrm{b}}(J_n)}{\hat{\sigma}_{\mathrm{b}}(n)} X_{J_n}^{(2)} + (1-F_\alpha)\frac{\hat{\sigma}_{\mathrm{w}}(J_n)}{\hat{\sigma}_{\mathrm{b}}(n)} Y_{J_n} + e'_{\mathrm{b}}(n),\\ Y_n &\stackrel{d}{=} \frac{\hat{\sigma}_{\mathrm{w}}(I_n)}{\hat{\sigma}_{\mathrm{w}}(n)} Y_{I_n}^{(1)} + F_\beta \frac{\hat{\sigma}_{\mathrm{w}}(J_n)}{\hat{\sigma}_{\mathrm{w}}(n)} Y_{J_n}^{(2)} + (1-F_\beta)\frac{\hat{\sigma}_{\mathrm{b}}(J_n)}{\hat{\sigma}_{\mathrm{w}}(n)} X_{J_n} + e'_{\mathrm{w}}(n),\end{aligned}$$

with conditions on independence and identical distributions analogous to (6.29)–(6.30). We have $\|e'_{\mathrm{b}}(n)\|_\infty, \|e'_{\mathrm{w}}(n)\|_\infty \to 0$, since the leading linear terms in the expansions (6.41)

https://doi.org/10.1017/S0963548314000364 Published online by Cambridge University Press

cancel out and the error terms $o(\sqrt{n})$ are asymptotically eliminated by the scaling of order $1/\sqrt{n}$. In view of (6.31) this suggests, for limits X and Y of X_n and Y_n, respectively,

$$X \stackrel{d}{=} \sqrt{U}X^{(1)} + F_\alpha\sqrt{1-U}X^{(2)} + (1-F_\alpha)\sqrt{1-U}Y^{(1)}, \tag{6.44}$$

$$Y \stackrel{d}{=} \sqrt{U}Y^{(1)} + F_\beta\sqrt{1-U}Y^{(2)} + (1-F_\beta)\sqrt{1-U}X^{(1)}, \tag{6.45}$$

where $X^{(1)}$, $X^{(2)}$, $Y^{(1)}$, $Y^{(2)}$ and U are independent and $X^{(1)}$, $X^{(2)}$ are distributed as X and $Y^{(1)}$, $Y^{(2)}$ are distributed as Y. We can apply Theorem 5.2 to the map associated to the system (6.44)–(6.45). The conditions (5.5) and (5.6) are trivially satisfied. Hence $(\mathcal{N}(0,1), \mathcal{N}(0,1))$ is the unique fixed point of the associated map in the space $\mathcal{M}_3^{\mathbb{R}}(0,1) \times \mathcal{M}_3^{\mathbb{R}}(0,1)$.

Theorem 6.6. *Consider the Pólya urn with random replacement matrix* (6.28) *with* $\alpha, \beta \in (0,1)$ *and* $\alpha + \beta < 3/2$ *and the normalized numbers* X_n *and* Y_n *of black balls as in* (6.43). *Then, as* $n \to \infty$,

$$X_n \xrightarrow{d} \mathcal{N}(0,1), \quad Y_n \xrightarrow{d} \mathcal{N}(0,1).$$

Proof. The proof is analogous to that of Theorem 6.2. □

Remark. The case $\alpha + \beta = 3/2$ differs in the error terms in (6.41) which then become $O(\sqrt{n})$. Since the variances in (6.42) get additional logarithmic factors we still obtain the system (6.44)–(6.45) and our proof technique still applies.

6.3. Cyclic urns

We fix an integer $m \geqslant 2$ and consider an urn with balls of types $1, \dots, m$. After a ball of type j is drawn, it is placed back into the urn together with a ball of type $j+1$ if $1 \leqslant j \leqslant m-1$ and together with a ball of type 1 if $j = m$. These urn models are called *cyclic urns*. Thus, the replacement matrix of a cyclic urn has the form

$$R = \begin{bmatrix} 0 & 1 & & & 0 \\ & 0 & 1 & & \\ & & 0 & & \\ & & & \ddots & 1 \\ 1 & & & & 0 \end{bmatrix}. \tag{6.46}$$

We let $R_n^{[j]}$ denote the number of type 1 balls after n draws when initially one ball of type j is contained in the urn. Our recursive approach described above yields the system of recursive distributional equations

$$\begin{aligned} R_n^{[1]} &\stackrel{d}{=} R_{I_n}^{[1]} + R_{J_n}^{[2]}, \\ R_n^{[2]} &\stackrel{d}{=} R_{I_n}^{[2]} + R_{J_n}^{[3]}, \\ &\vdots \\ R_n^{[m]} &\stackrel{d}{=} R_{I_n}^{[m]} + R_{J_n}^{[1]}, \end{aligned} \tag{6.47}$$

where, on the right-hand sides, I_n and $R_k^{[j]}$ for $j = 1, \ldots, m$, $k = 0, \ldots, n-1$ are independent, I_n uniformly distributed on $\{0, \ldots, n-1\}$ and $J_n = n - 1 - I_n$.

We denote the imaginary unit by i and use the primitive roots of unity

$$\omega := \omega_m := \exp\left(\frac{2\pi \mathrm{i}}{m}\right) =: \lambda + \mathrm{i}\mu \tag{6.48}$$

with $\lambda, \mu \in \mathbb{R}$. Note that for $2 \leqslant m \leqslant 6$ we have $\lambda \leqslant 1/2$, while for $m \geqslant 7$ we have $\lambda > 1/2$. Asymptotic expressions for the mean of the $R_n^{[j]}$ can be found (together with further analysis) in [15, 16, 31]. To keep this section self-contained we give an exact formula for later use.

Lemma 6.7. *Let $R_n^{[j]}$ be the number of balls of colour 1 after n draws in a cyclic urn with $m \geqslant 2$ colours, starting with one ball of colour j. Then, with $\omega = \omega_m$ as in (6.48) we have*

$$\mathbb{E}\big[R_n^{[j]}\big] = \frac{n+1}{m} + \frac{1}{m} \sum_{k \in \{1,\ldots,m-1\} \setminus \{m/2\}} \frac{\Gamma(n+1+\omega^k)}{\Gamma(n+1)\,\Gamma(\omega^k+1)} \omega^{k(j-1)}. \tag{6.49}$$

In particular, we have $\mathbb{E}[R_n^{[j]}] = \frac{1}{m} n + \mathrm{O}(1)$ for $m = 2, 3, 4$ and, for $m > 4$, as $n \to \infty$,

$$\mathbb{E}\big[R_n^{[j]}\big] = \frac{1}{m} n + \Re(\kappa_j n^{\mathrm{i}\mu}) n^{\lambda} + o(n^{\lambda}), \qquad \kappa_j := \frac{2\omega^{j-1}}{m\Gamma(\omega+1)}. \tag{6.50}$$

Proof. Using the system (6.47), we obtain by conditioning on I_n, for any $1 \leqslant j \leqslant m$,

$$\begin{aligned}
\mathbb{E}\big[R_n^{[j]}\big] &= \frac{1}{n} \sum_{i=0}^{n-1} \mathbb{E}\big[R_i^{[j]}\big] + \frac{1}{n} \sum_{i=0}^{n-1} \mathbb{E}\big[R_i^{[j+1]}\big] \\
&= \frac{1}{n}\big(\mathbb{E}\big[R_{n-1}^{[j]}\big] + \mathbb{E}\big[R_{n-1}^{[j+1]}\big]\big) + \frac{n-1}{n} \mathbb{E}\big[R_{n-1}^{[j]}\big] \\
&= \mathbb{E}\big[R_{n-1}^{[j]}\big] + \frac{1}{n} \mathbb{E}\big[R_{n-1}^{[j+1]}\big],
\end{aligned}$$

where we set $R_i^{[m+1]} := R_i^{[1]}$ for any $1 \leqslant i \leqslant n$. With column vector $R_n := (R_n^{[1]}, \ldots, R_n^{[m]})$, the replacement matrix R in (6.46) and the identity matrix Id_m, this is rewritten as

$$\mathbb{E}\big[R_n\big] = \left(\mathrm{Id}_m + \frac{1}{n} R\right) \mathbb{E}\big[R_{n-1}\big] = \prod_{k=1}^{n} \left(\mathrm{Id}_m + \frac{1}{k} R\right) \mathbb{E}\big[R_0\big].$$

The eigenvalues of the replacement matrix are all mth roots of unity ω^k, $k = 1, \ldots, m$, and a possible eigenbasis is $v_k := \frac{1}{m}(\omega^0, \omega^k, \ldots, \omega^{(m-1)k})^t$, $k = 1, \ldots, m$. Decomposing the mapping induced by R into the projections π_{v_k} onto the respective eigenspaces, we obtain

$$\begin{aligned}
\prod_{\ell=1}^{n} \left(\mathrm{Id}_m + \frac{1}{\ell} R\right) &= \sum_{k=1}^{m} \prod_{\ell=1}^{n} \left(1 + \frac{1}{\ell} \omega^k\right) \pi_{v_k} \\
&= (n+1)\, \pi_{v_m} + \sum_{k \in \{1,\ldots,m-1\} \setminus \{m/2\}} \frac{\Gamma(n+1+\omega^k)}{\Gamma(\omega^k+1)\Gamma(n+1)} \pi_{v_k}.
\end{aligned}$$

Moreover, $\pi_{v_k}(\mathbb{E}[R_0]) = v_k$ and $v_m = \frac{1}{m}(1,\ldots,1)$, hence the jth component of the latter display implies (6.49). The asymptotic expansion in (6.50) is now directly read off: note that the roots of unity come in conjugate pairs $\omega^{m-k} = \overline{\omega}^k$. If m is even, $\omega^{m/2} = \overline{\omega}^{m/2} = -1$, otherwise only $\omega^m = 1$ is real. Combining pairs of summands for such conjugate pairs and using $\Gamma(\overline{z}) = \overline{\Gamma(z)}$, we obtain the terms

$$\frac{\Gamma(n+1+\omega^k)\,\omega^{(j-1)k}}{\Gamma(n+1)\,\Gamma(\omega^k+1)} + \frac{\Gamma(n+1+\overline{\omega}^k)\overline{\omega}^{(j-1)k}}{\Gamma(n+1)\,\Gamma(\overline{\omega}^k+1)} = 2\,\Re\left(\frac{\omega^{(j-1)k}\,\Gamma(n+1+\omega^k)}{\Gamma(\omega^k+1)\,\Gamma(n+1)}\right).$$

By Stirling approximation the asymptotic growth order of the latter term is $\Re(n^{\omega^k})$, hence the dominant asymptotic term is for the conjugate pair with largest real part, ω and ω^{m-1}. This implies (6.50) for $m > 4$. For $m = 3, 4$ the periodic term is $o(1)$, respectively $\mathrm{O}(1)$; for $m = 2$ there is no periodic fluctuation. □

We do not discuss limit laws for the cases $2 \leqslant m \leqslant 6$ in detail. They lead to asymptotic normality, as has been shown with different proofs by Janson [15] and [16, Example 7.9]. These cases can be covered by our approach similarly to the normal cases in Sections 6.1 and 6.2. For $2 \leqslant m \leqslant 6$, the system of limit equations is

$$\begin{aligned} X^{[1]} &\stackrel{d}{=} \sqrt{U}X^{[1]} + \sqrt{1-U}X^{[2]}, \\ X^{[2]} &\stackrel{d}{=} \sqrt{U}X^{[2]} + \sqrt{1-U}X^{[3]}, \\ &\vdots \\ X^{[m]} &\stackrel{d}{=} \sqrt{U}X^{[m]} + \sqrt{1-U}X^{[1]}, \end{aligned}$$

and Theorem 5.2 applies.

We now assume $m \geqslant 7$. In particular, we have the asymptotic expansion (6.50) of the mean of the $R_n^{[j]}$ with $\lambda > 1/2$. We define the normalizations

$$X_n^{[j]} := \frac{R_n^{[j]} - \frac{1}{m}n}{n^\lambda}. \tag{6.51}$$

Hence, we obtain for the $X_n^{[j]}$ the system

$$\begin{aligned} X_n^{[1]} &\stackrel{d}{=} \left(\frac{I_n}{n}\right)^\lambda X_{I_n}^{[1]} + \left(\frac{J_n}{n}\right)^\lambda X_{J_n}^{[2]} - \frac{1}{mn^\lambda}, \\ X_n^{[2]} &\stackrel{d}{=} \left(\frac{I_n}{n}\right)^\lambda X_{I_n}^{[2]} + \left(\frac{J_n}{n}\right)^\lambda X_{J_n}^{[3]} - \frac{1}{mn^\lambda}, \\ &\vdots \\ X_n^{[m]} &\stackrel{d}{=} \left(\frac{I_n}{n}\right)^\lambda X_{I_n}^{[m]} + \left(\frac{J_n}{n}\right)^\lambda X_{J_n}^{[1]} - \frac{1}{mn^\lambda}, \end{aligned}$$

where, on the right-hand sides, I_n and $X_k^{[j]}$ for $j = 1,\ldots,m$, $k = 0,\ldots,n-1$ are independent. To describe the asymptotic periodic behaviour of the distributions of the $X_n^{[j]}$, we use the following related system of limit equations:

$$
\begin{aligned}
X^{[1]} &\stackrel{d}{=} U^\omega X^{[1]} + (1-U)^\omega X^{[2]},\\
X^{[2]} &\stackrel{d}{=} U^\omega X^{[2]} + (1-U)^\omega X^{[3]},\\
&\vdots\\
X^{[m]} &\stackrel{d}{=} U^\omega X^{[m]} + (1-U)^\omega X^{[1]}.
\end{aligned}
$$

Since ω is complex non-real, this now has to be considered as a system to solve for distributions $\mathcal{L}(X^{[1]}),\dots,\mathcal{L}(X^{[m]})$ on the complex plane $\mathbb{C}$. The corresponding map $\bar{T}$ is a special case of T' in (5.3):

$$
\begin{aligned}
\bar{T} &: \mathcal{M}^{\mathbb{C},\times m} \to \mathcal{M}^{\mathbb{C},\times m},\\
&(\mu_1,\dots,\mu_m) \mapsto (\bar{T}_1(\mu_1,\dots,\mu_m),\dots,\bar{T}_m(\mu_1,\dots,\mu_m)),\\
\bar{T}_j(\mu_1,\dots,\mu_m) &:= \mathcal{L}\big(U^\omega V^{[j]} + (1-U)^\omega V^{[j+1]}\big)
\end{aligned}
\tag{6.52}
$$

for $j=1,\dots,m$, where $U, V^{[1]},\dots,V^{[m+1]}$ are independent, U is uniformly distributed on $[0,1]$ and $\mathcal{L}(V^{[j]})=\mu_j$ for $j=1,\dots,m$ and $\mathcal{L}(V^{[m+1]})=\mu_1$.

Lemma 6.8. *Let $m \geqslant 7$. The restriction of $\bar{T}$ to $\mathcal{M}_2^{\mathbb{C}}(\kappa_1)\times\cdots\times\mathcal{M}_2^{\mathbb{C}}(\kappa_m)$ has a unique fixed point.*

Proof. We verify the conditions of Theorem 5.3. First note that condition (5.7) for our $\bar{T}$ in (6.52) is

$$
\mathbb{E}\big[U^\omega\big]\kappa_j + \mathbb{E}\big[(1-U)^\omega\big]\kappa_{j+1} = \kappa_j, \quad j=1,\dots,m, \tag{6.53}
$$

with $\kappa_{m+1} := \kappa_1$. Since

$$
\mathbb{E}[U^\omega] = \mathbb{E}[(1-U)^\omega] = (1+\omega)^{-1}
$$

and $\kappa_{j+1} = \omega\kappa_j$, we find that (6.53) is satisfied. Condition (5.8) for our $\bar{T}$ is

$$
\mathbb{E}\big[|U^{2\omega}|\big] + \mathbb{E}\big[|(1-U)^{2\omega}|\big] < 1.
$$

Since $m \geqslant 7$, we have $\lambda > 1/2$, and thus

$$
\mathbb{E}[|U^{2\omega}|] + \mathbb{E}[|(1-U)^{2\omega}|] = 2/(1+2\lambda) < 1.
$$

Hence Theorem 5.3 applies, and implies the assertion. □

The fixed point in Lemma 6.8 has a particularly simple structure, as follows. Note that a description related to (6.54) was given in Remark 2.3 in Janson [18].

Lemma 6.9. *Let $m \geqslant 7$ and $(\mathcal{L}(\Lambda^{[1]}),\dots,\mathcal{L}(\Lambda^{[m]}))$ be the unique fixed point in Lemma 6.8. Furthermore, let $\mathcal{L}(\Lambda)$ be the (unique) fixed point of*

$$
X \stackrel{d}{=} U^\omega X + \omega(1-U)^\omega X' \quad \text{in } \mathcal{M}_2^{\mathbb{C}}\left(\frac{2}{m\Gamma(\omega+1)}\right), \tag{6.54}
$$

https://doi.org/10.1017/S0963548314000364 Published online by Cambridge University Press

where X, X' and U are independent, U is uniformly distributed on $[0,1]$, *and X and X' have identical distributions. Then we have*

$$\Lambda^{[j]} \stackrel{d}{=} \omega^{j-1}\Lambda, \quad j = 1,\dots,m.$$

Proof. We abbreviate $\gamma := 2/(m\Gamma(\omega+1))$. For X, X' and U independent, U uniformly distributed on $[0,1]$, and X and X' identically distributed with $\mathbb{E}X = \gamma$, we have

$$\mathbb{E}\big[U^\omega X + \omega(1-U)^\omega X'\big] = \frac{1}{1+\omega}(\gamma + \omega\gamma) = \gamma,$$

hence the map of probability measures on $\mathbb{C}$ associated to (6.54) maps $\mathcal{M}_2^{\mathbb{C}}(\gamma)$ into itself. The argument of the proof of Theorem 5.3 implies that this map is a contraction on $(\mathcal{M}_2^{\mathbb{C}}(\gamma), \ell_2)$. Hence it has a unique fixed point $\mathcal{L}(\Lambda)$. We have

$$(\mathcal{L}(\Lambda), \mathcal{L}(\omega\Lambda), \dots, \mathcal{L}(\omega^{m-1}\Lambda)) \in \mathcal{M}_2^{\mathbb{C}}(\kappa_1) \times \cdots \times \mathcal{M}_2^{\mathbb{C}}(\kappa_m)$$

and, by plugging into (6.52), we find that this vector is a fixed point of $\bar{T}$. Since, by Lemma 6.8, there is only one fixed point of $\bar{T}$ in $\mathcal{M}_2^{\mathbb{C}}(\kappa_1) \times \cdots \times \mathcal{M}_2^{\mathbb{C}}(\kappa_m)$, the assertion follows. □

The asymptotic periodic behaviour in the following theorem has already been shown almost surely by martingale methods in [31, Section 4.2]; see also [16, Theorem 3.24]. Our contraction approach adds the characterization of $\mathcal{L}(\Lambda)$ as the fixed point in (6.54). The proof is based on the complex version of the ℓ_2-metric and resembles ideas from Fill and Kapur [12]; see also [21, Theorem 5.3].

Theorem 6.10. *Let $m \geqslant 7$ and $X_n^{[j]}$ be as in* (6.51) *and let $\mathcal{L}(\Lambda)$ be the unique fixed point in Lemma 6.9. Then, for all $j = 1,\dots,m$, we have*

$$\ell_2\big(X_n^{[j]}, \Re(\mathrm{e}^{\mathrm{i}(\mu\ln(n) + 2\pi\frac{j-1}{m})}\Lambda)\big) \to 0 \quad (n \to \infty). \tag{6.55}$$

Proof. Let $\Lambda^{[1]},\dots,\Lambda^{[m]}$ be independent random variables such that $(\mathcal{L}(\Lambda^{[1]}),\dots,\mathcal{L}(\Lambda^{[m]}))$ is the unique fixed point as in Lemma 6.8. Set $\Lambda^{[m+1]} := \Lambda^{[1]}$. Note that for the random variable within the real part in (6.55) with Lemma 6.9, we have

$$\mathrm{e}^{\mathrm{i}(\mu\ln(n) + 2\pi\frac{j-1}{m})}\Lambda = n^{\mathrm{i}\mu}\omega^{j-1}\Lambda \stackrel{d}{=} n^{\mathrm{i}\mu}\Lambda^{[j]}.$$

The fixed point property of the $\Lambda^{[j]}$ implies

$$\Re\big(n^{\mathrm{i}\mu}\Lambda^{[j]}\big) \stackrel{d}{=} \Re\big(n^{\mathrm{i}\mu}U^\omega\Lambda^{[j]}\big) + \Re\big(n^{\mathrm{i}\mu}(1-U)^\omega\Lambda^{[j+1]}\big)$$

for all $j = 1,\dots,m$ and $n \geqslant 0$. We denote

$$\Delta_j(n) := \ell_2\big(X_n^{[j]}, \Re(n^{\mathrm{i}\mu}\Lambda^{[j]})\big)$$

and set $\Delta_{m+1}(n) := \Delta_1(n)$. Now, we assume that the $X_n^{[j]}$, $\Lambda^{[j]}$, $n \geqslant 1$, $1 \leqslant j \leqslant m$, I_n, U appearing in (6.51) and (6.52) are defined on one probability space such that

https://doi.org/10.1017/S0963548314000364 Published online by Cambridge University Press

$(X_n^{[j]}, \Re(n^{i\mu}\Lambda^{[j]}))$ are optimal ℓ_2-couplings for all $n \geqslant 0$ and all $1 \leqslant j \leqslant m$ and such that $I_n = \lfloor nU \rfloor$. Then we have

$$\begin{aligned}
&\Delta_j(n) \\
&= \ell_2\left(\left(\frac{I_n}{n}\right)^{\lambda} X_{I_n}^{[j]} + \left(\frac{J_n}{n}\right)^{\lambda} X_{J_n}^{[j+1]} - \frac{1}{mn^{\lambda}}, \Re\big(n^{i\mu}U^{\omega}\Lambda^{[j]}\big) + \Re\big(n^{i\mu}(1-U)^{\omega}\Lambda^{[j+1]}\big)\right) \\
&\leqslant \left\| \left\{ \left(\frac{I_n}{n}\right)^{\lambda} X_{I_n}^{[j]} - \Re\left(\frac{I_n^{\omega}}{n^{\lambda}}\Lambda^{[j]}\right) \right\} + \left\{ \left(\frac{J_n}{n}\right)^{\lambda} X_{J_n}^{[j+1]} - \Re\left(\frac{J_n^{\omega}}{n^{\lambda}}\Lambda^{[j+1]}\right) \right\} \right\|_2 \\
&\quad + \left\| \Re\left(\frac{I_n^{\omega}}{n^{\lambda}}\Lambda^{[j]}\right) - \Re\big(n^{i\mu}U^{\omega}\Lambda^{[j]}\big) \right\|_2 + \left\| \Re\left(\frac{J_n^{\omega}}{n^{\lambda}}\Lambda^{[j+1]}\right) - \Re\big(n^{i\mu}U^{\omega}\Lambda^{[j+1]}\big) \right\|_2 + \frac{1}{mn^{\lambda}} \\
&=: S_1 + S_2 + S_3 + \frac{1}{mn^{\lambda}}.
\end{aligned} \tag{6.56}$$

First note that the summands S_2 and S_3 tend to zero. We have $(I_n/n)^{\omega} \to U^{\omega}$ almost surely by $I_n = \lfloor nU \rfloor$. Since $\Lambda^{[j]}$ and $\Lambda^{[j+1]}$ have finite second moments, we can apply dominated convergence to obtain $S_2, S_3 \to 0$ as $n \to \infty$.

For the estimate of the first summand S_1, we abbreviate

$$W_n^{[j]} := \left(\frac{I_n}{n}\right)^{\lambda} X_{I_n}^{[j]} - \Re\left(\frac{I_n^{\omega}}{n^{\lambda}}\Lambda^{[j]}\right), \qquad W_n^{[j+1]} := \left(\frac{J_n}{n}\right)^{\lambda} X_{J_n}^{[j+1]} - \Re\left(\frac{J_n^{\omega}}{n^{\lambda}}\Lambda^{[j+1]}\right).$$

Then we have

$$S_1^2 = \mathbb{E}\big[(W_n^{[j]})^2\big] + \mathbb{E}\big[(W_n^{[j+1]})^2\big] + 2\mathbb{E}\big[W_n^{[j]}W_n^{[j+1]}\big]. \tag{6.57}$$

Conditioning on I_n and using that $(X_k^{[j]}, \Re(k^{i\mu}\Lambda^{[j]}))$ are optimal ℓ_2-couplings, we obtain

$$\begin{aligned}
\mathbb{E}\big[(W_n^{[j]})^2\big] &= \sum_{k=0}^{n-1} \frac{1}{n}\mathbb{E}\left[\left\{\left(\frac{k}{n}\right)^{\lambda} X_k^{[j]} - \Re\left(\frac{k^{\lambda}k^{i\mu}}{n^{\lambda}}\Lambda^{[j]}\right)\right\}^2\right] \\
&= \sum_{k=0}^{n-1} \frac{1}{n}\left(\frac{k}{n}\right)^{2\lambda} \mathbb{E}\big[\{X_k^{[j]} - \Re\big(k^{i\mu}\Lambda^{[j]}\big)\}^2\big] \\
&= \sum_{k=0}^{n-1} \frac{1}{n}\left(\frac{k}{n}\right)^{2\lambda} \Delta_j^2(k) \\
&= \mathbb{E}\left[\left(\frac{I_n}{n}\right)^{2\lambda} \Delta_j^2(I_n)\right].
\end{aligned}$$

Analogously, we have

$$\mathbb{E}[(W_n^{[j+1]})^2] = \mathbb{E}\left[\left(\frac{J_n}{n}\right)^{2\lambda} \Delta_{j+1}^2(J_n)\right].$$

To bound the mixed term in (6.57), note that by the expansion (6.50) and the normalization (6.51) we have $\mathbb{E}[X_n^{[j]}] = \Re(\kappa_j n^{i\mu}) + r_j(n)$ with $r_j(n) \to 0$ as $n \to \infty$ for all $j = 1, \ldots, m$. In particular, we have $\|r_j\|_{\infty} < \infty$. Together with $\mathbb{E}[\Lambda^{[j]}] = \kappa_j$, this implies

$$\mathbb{E}[W_n^{[j]}] = \mathbb{E}[(I_n/n)^{\lambda} r_j(I_n)]$$

https://doi.org/10.1017/S0963548314000364 Published online by Cambridge University Press

and

$$\mathbb{E}\big[W_n^{[j]} W_n^{[j+1]}\big] = \mathbb{E}\left[\left(\frac{I_n}{n}\frac{J_n}{n}\right)^{\lambda} r_j(I_n) r_{j+1}(J_n)\right]. \tag{6.58}$$

To show that the latter term tends to zero, let $\varepsilon > 0$. Then there exists $k_0 \in \mathbb{N}$ such that $r_j(k) < \varepsilon$, $r_{j+1}(k) < \varepsilon$ for all $k \geqslant k_0$. For all $n > 2k_0$ we obtain, by considering the event

$$\{k_0 \leqslant I_n \leqslant n - 1 - k_0\}$$

and its complement,

$$\mathbb{E}\big[W_n^{[j]} W_n^{[j+1]}\big] \leqslant \frac{2k_0}{n} \|r_j\|_\infty \|r_{j+1}\|_\infty + \varepsilon^2.$$

Hence, we obtain that the mixed term (6.58) tends to zero.

Altogether, we obtain from (6.56) as $n \to \infty$ that

$$\begin{aligned}\Delta_j(n) &\leqslant \left\{\mathbb{E}\left[\left(\frac{I_n}{n}\right)^{2\lambda} \Delta_j^2(I_n)\right] + \mathbb{E}\left[\left(\frac{J_n}{n}\right)^{2\lambda} \Delta_{j+1}^2(J_n)\right] + o(1)\right\}^{1/2} + o(1)\\ &\leqslant \left\{2\mathbb{E}\left[\left(\frac{I_n}{n}\right)^{2\lambda} \Delta^2(I_n)\right] + o(1)\right\}^{1/2} + o(1),\end{aligned}$$

for all $j = 1, \ldots, m$, where

$$\Delta(n) := \max_{1 \leqslant j \leqslant m} \Delta_j(n).$$

Hence, we have

$$\Delta(n) \leqslant \left\{2\mathbb{E}\left[\left(\frac{I_n}{n}\right)^{2\lambda} \Delta^2(I_n)\right] + o(1)\right\}^{1/2} + o(1). \tag{6.59}$$

Now, we obtain $\Delta(n) \to 0$ as in the proof of Theorem 6.1. First from (6.59) we obtain with $I_n/n \to U$ almost surely that

$$\begin{aligned}\Delta(n) &\leqslant \left\{2\mathbb{E}\left[\left(\frac{I_n}{n}\right)^{2\lambda}\right] \max_{0 \leqslant k \leqslant n-1} \Delta^2(k) + o(1)\right\}^{1/2} + o(1)\\ &\leqslant \left\{\left(\frac{2}{1+2\lambda} + o(1)\right) \max_{0 \leqslant k \leqslant n-1} \Delta^2(k) + o(1)\right\}^{1/2} + o(1).\end{aligned}$$

Since $\lambda > 1/2$ this implies that the sequence $(\Delta(n))_{n \geqslant 0}$ is bounded. We set $\eta := \sup_{n \geqslant 0} \Delta(n)$ and $\xi := \limsup_{n \to \infty} \Delta(n)$. For any $\varepsilon > 0$ there exists an $n_0 \geqslant 0$ such that $\Delta(n) \leqslant \xi + \varepsilon$ for all $n \geqslant n_0$. Hence, from (6.59) we obtain

$$\Delta(n) \leqslant \left\{2\mathbb{E}\left[\mathbf{1}_{\{I_n < n_0\}}\left(\frac{I_n}{n}\right)^{2\lambda}\right]\eta^2 + 2\mathbb{E}\left[\mathbf{1}_{\{I_n \geqslant n_0\}}\left(\frac{I_n}{n}\right)^{2\lambda}\right](\xi + \varepsilon)^2 + o(1)\right\}^{1/2} + o(1).$$

With $n \to \infty$ this implies

$$\xi \leqslant \sqrt{\frac{2}{1+2\lambda}}(\xi + \varepsilon).$$

Since $\sqrt{2/(1+2\lambda)} < 1$ and $\varepsilon > 0$ is arbitrary, this implies $\xi = 0$. □

https://doi.org/10.1017/S0963548314000364 Published online by Cambridge University Press

7. Remarks on the use of the contraction method

A novel technical aspect of this paper is that we extend the use of the contraction method to systems of recursive distributional equations. Alternatively, one may be tempted to couple the random variables B_n^{b} and B_n^{w} in (2.2) and (2.3) on one probability space, set up a recurrence for their vector $(B_n^{\mathrm{b}}, B_n^{\mathrm{w}})$ and try to apply general transfer theorems from the contraction method for multivariate recurrences, such as Theorem 4.1 in Neininger [28] or Theorem 4.1 in Neininger and Rüschendorf [29]. For some particular instances (replacement schemes) of the Pólya urn this is in fact possible. However, when attempting to come up with a limit theory of the generality of the present paper, such a multivariate approach hits two snags that seem difficult to overcome. In this section we highlight these problems using one of the examples discussed above, and explain why we consider such a multivariate approach disadvantageous in the context of Pólya urns.

We consider the example from Section 6.2 with the random replacement matrix in (6.28) and denote the bivariate random variable by $B_n := (B_n^{\mathrm{b}}, B_n^{\mathrm{w}})$ with B_n^{b} and B_n^{w} as in (6.29) and (6.30) respectively. Note that in the discussion of Section 6.2 the random variables B_n^{b} and B_n^{w} did not need to be defined on a common probability space. Hence, first of all, only the marginals of B_n are determined by the urn process, and we have the choice of a joint distribution for B_n respecting these marginals. We could keep the components independent or choose appropriate couplings. We choose a form that implies a recurrence of the form typically considered in general limit theorems from the contraction method. The coupling is defined recursively by $B_0 = (1,0)$ and, for $n \geqslant 1$,

$$B_n :\stackrel{d}{=} B_{I_n} + \begin{bmatrix} F_\alpha & 1-F_\alpha \\ 1-F_\beta & F_\beta \end{bmatrix} B'_{J_n}, \tag{7.1}$$

where $(B_n)_{0\leqslant k<n}$, $(B'_n)_{0\leqslant k<n}$, (F_α, F_β), and I_n are independent and B_k and B'_k identically distributed for all $0 \leqslant k < n$. As in Section 6.2, I_n is uniformly distributed on $\{0,\dots,n-1\}$ and $J_n := n-1-I_n$, while F_α and F_β are Bernoulli random variables, being 1 with probabilities α and β respectively, and otherwise 0. Note that for any joint distribution of (F_α, F_β), definition (7.1) leads to a sequence $(B_n)_{n\geqslant 1}$ with correct marginals of B_n^{b} and B_n^{w}. A beneficial joint distribution of (F_α, F_β) will be chosen below.

We consider the cases where $\alpha+\beta-1 < 1/2$. Since these lead to normal limits, one may try to apply Theorem 4.1 in [29], where $2 < s \leqslant 3$ is the index of the Zolotarev metric ζ_s on which that theorem is based. The best possible contraction condition (see [29, equation (25)]) is obtained with $s = 3$, which we fix subsequently. Now, for the application of Theorem 4.1 in [29] we need an asymptotic expansion of the covariance matrix of B_n. In view of Lemma 6.5, we assume that for all $i, j = 1, 2$ we have

$$(\mathrm{Cov}(B_n))_{ij} = f_{ij} n + o(n), \quad (n \to \infty) \tag{7.2}$$

such that $(f_{ij})_{ij}$ is a symmetric, positive definite 2×2 matrix. Hence there exists an $n_1 \geqslant 1$ such that $\mathrm{Cov}(B_n)$ is positive definite for all $n \geqslant n_1$. For the normalized random sequence

$$X_n := (\mathrm{Cov}(B_n))^{-1/2}(B_n - \mathbb{E}[B_n]), \quad n \geqslant n_1,$$

https://doi.org/10.1017/S0963548314000364 Published online by Cambridge University Press

we obtain the limit equation

$$X \stackrel{d}{=} \sqrt{U} X + \sqrt{1-U} \begin{bmatrix} F_\alpha & 1-F_\alpha \\ 1-F_\beta & F_\beta \end{bmatrix} X',$$

where $X, X', U, (F_\alpha, F_\beta)$ are independent, X and X' are identically distributed and U is uniformly distributed on $[0,1]$. Now the application of Theorem 4.1 in [29] requires condition (25) there to be satisfied, which in our example is written as

$$\mathbb{E}\big[U^{3/2}\big] + \mathbb{E}\big[(1-U)^{3/2}\big]\mathbb{E}\left[\left\| \begin{bmatrix} F_\alpha & 1-F_\alpha \\ 1-F_\beta & F_\beta \end{bmatrix} \right\|_{\mathrm{op}}^3\right] < 1, \tag{7.3}$$

where $\|\cdot\|_{\mathrm{op}}$ denotes the operator norm of the matrix. Here, the joint distribution of (F_α, F_β) can be chosen to minimize the left-hand side of the latter inequality as follows. For V uniformly distributed and independent of U, we set $F_\alpha = \mathbf{1}_{\{V \leqslant \alpha\}}$ and $F_\beta = \mathbf{1}_{\{V \leqslant \beta\}}$. With this choice of the joint distribution of (F_α, F_β), condition (7.3) turns into

$$\frac{2}{5}\big(2 + |\alpha - \beta|(2^{3/2} - 1)\big) < 1.$$

We see that this condition is not satisfied in the whole range $\alpha + \beta - 1 < 1/2$. Hence, in the best possible setup that we could find, Theorem 4.1 in [29] does not yield results of the strength of Theorem 6.6.

A second drawback of the use of multivariate recurrences is that we needed the assumption of the expansion (7.2), which is technically required in order to verify condition (24) in [29]. Hence, after coupling B_n^{b} and B_n^{w} on one probability space such that we may satisfy (7.3), we have to derive asymptotic expressions for the covariance $\mathrm{Cov}(B_n^{\mathrm{b}}, B_n^{\mathrm{w}})$ and to identify the leading constant in these asymptotics. Note that this covariance is meaningless for the Pólya urn and only emerges by artificially coupling B_n^{b} and B_n^{w}. This covariance does not appear in the approach we propose in Section 6, which makes its application much simpler compared to a multivariate formulation.

A reason why our approach of analysing systems of recurrences is more powerful than the use of multivariate recurrences is found when comparing the spaces of probability measures with the aim of applying contraction arguments to them. In Section 4 we introduce the space $(\mathcal{M}_s^{\mathbb{R}})^{\times d}$ in (4.1) and work on subspaces where first, or first and second, moments of the probability measures are fixed. The corresponding space in a multivariate formulation and in Theorem 4.1 in [29] is the space $\mathcal{M}_s(\mathbb{R}^d)$ of all probability measures on $\mathbb{R}^d$ with finite absolute sth moment. Clearly $(\mathcal{M}_s^{\mathbb{R}})^{\times d}$ is much smaller than $\mathcal{M}_s(\mathbb{R}^d)$, *e.g.*, the first space can be embedded into the second by forming product measures. This makes it plausible that it is much easier to find contracting maps as developed in Section 5 on $(\mathcal{M}_s^{\mathbb{R}})^{\times d}$ than on $\mathcal{M}_s(\mathbb{R}^d)$, and we feel that this causes the problems mentioned above with a multivariate formulation.

In the dissertation by Knape [23, Chapter 5], more details of our use of the contraction method and an alternative multivariate formulation are given. There, too, improved versions of Theorem 4.1 in [29] are derived by a change of the underlying probability metric, which lead to better conditions compared to (7.3). However, the need to derive artificial covariances in a multivariate approach, as discussed above, could not be surmounted in [23]. Similar advantages of the use of systems of recurrences over multivariate formulations were noted in Leckey, Neininger and Szpankowski [25, Section 7].

Acknowledgements

We thank two referees for their comments and careful reading. We also thank the e-print archive `arXiv.org` and Cornell University Library for making an electronic preprint of this work freely and publicly available by 16 January 2013.

References

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

[2] 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.

[3] Bagchi, A. and Pal, A. K. (1985) Asymptotic normality in the generalized Polya–Eggenberger urn model, with an application to computer data structures. *SIAM J. Algebraic Discrete Methods* **6** 394–405.

[4] Bai, Z. D. and Hu, F. (1999) Asymptotic theorems for urn models with nonhomogeneous generating matrices. *Stochastic Process. Appl.* **80** 87–101.

[5] Bai, Z. D., Hu, F. and Zhang, L.-X. (2002) Gaussian approximation theorems for urn models and their applications. *Ann. Appl. Probab.* **12** 1149–1173.

[6] Bickel, P. J. and Freedman, D. A. (1981) Some asymptotic theory for the bootstrap. *Ann. Statist.* **9** 1196–1217.

[7] Chauvin, B., Liu, Q. and Pouyanne, N. (2012) Support and density of the limit m-ary search tree distribution. In *23rd International Meeting on Probabilistic, Combinatorial, and Asymptotic Methods for the Analysis of Algorithms: AofA'12*, DMTCS Proc. **AQ** 2012, pp. 191–200.

[8] Chauvin, B., Liu, Q. and Pouyanne, N. (2014) Limit distributions for multitype branching processes of m-ary search trees. *Ann. IHP.* **50** 628–654.

[9] Chauvin, B., Mailler, C. and Pouyanne, N. (2013) Smoothing equations for large Pólya urns. `arXiv:1302.1412`

[10] Chauvin, B., Pouyanne, N. and Sahnoun, R. (2011) Limit distributions for large Pólya urns. *Ann. Appl. Probab.* **21** 1–32.

[11] Drmota, M., Janson, S. and Neininger, R. (2008) A functional limit theorem for the profile of search trees. *Ann. Appl. Probab.* **18** 288–333.

[12] Fill, J. A. and Kapur, N. (2004) The space requirement of m-ary search trees: Distributional asymptotics for $m \geqslant 27$. Invited paper, *Proc. 7th Iranian Statistical Conference, 2004.* `www.ams.jhu.edu/~fill/papers/periodic.pdf`

[13] Flajolet, P., Gabarró, J. and Pekari, H. (2005) Analytic urns. *Ann. Probab.* **33** 1200–1233.

[14] Hwang, H.-K. and Neininger, R. (2002) Phase change of limit laws in the quicksort recurrence under varying toll functions. *SIAM J. Comput.* **31** 1687–1722.

[15] Janson, S. (1983) Limit theorems for certain branching random walks on compact groups and homogeneous spaces. *Ann. Probab.* **11** 909–930.

[16] Janson, S. (2004) Functional limit theorem for multitype branching processes and generalized Pólya urns. *Stochastic Process. Appl.* **110** 177–245.

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

[18] Janson, S. (2006) Congruence properties of depths in some random trees. *Alea* **1** 347–366.

[19] Janson, S. (2010) Moments of gamma type and the Brownian supremum process area. *Probab. Surv.* **7** 1–52.

[20] Janson, S. and Kaijser, S. (2012) Higher moments of Banach space valued random variables. *Mem. Amer. Math. Soc.*, to appear.
[21] Janson, S. and Neininger, R. (2008) The size of random fragmentation trees. *Probab. Theory Rel. Fields* **142** 399–442.
[22] Johnson, N. L. and Kotz, S. (1977) *Urn Models and their Application: An Approach to Modern Discrete Probability Theory*, Wiley Series in Probability and Mathematical Statistics, Wiley.
[23] Knape, M. (2013) Pólya urns via the contraction method. PhD dissertation. Submitted at the J.W. Goethe University, Frankfurt am Main, April 2013. urn:nbn:de:hebis:30:3-322846
[24] Kotz, S., Mahmoud, H. M. and Robert, P. (2000) On generalized Pólya urn models. *Statist. Probab. Lett.* **49** 163–173.
[25] Leckey, K., Neininger, R. and Szpankowski, W. (2013) Towards more realistic probabilistic models for data structures: The external path length in tries under the Markov model. In *Proc. ACM–SIAM Symposium on Discrete Algorithms* (*SODA*), pp. 877–886.
[26] Mahmoud, H. M. (2009) *Pólya Urn Models*, Texts in Statistical Science Series, CRC Press.
[27] Matthews, P. C. and Rosenberger, W. F. (1997) Variance in randomized play-the-winner clinical trials. *Statist. Probab. Lett.* **35** 233–240.
[28] Neininger, R. (2001) On a multivariate contraction method for random recursive structures with applications to quicksort. *Random Struct. Alg.* **19** 498–524.
[29] Neininger, R. and Rüschendorf, L. (2004) A general limit theorem for recursive algorithms and combinatorial structures. *Ann. Appl. Probab.* **14** 378–418.
[30] Neininger, R. and Sulzbach, H. (2012) On a functional contraction method. *Ann. Probab.*, to appear. arXiv:1202.1370
[31] Pouyanne, N. (2005) Classification of large Pólya–Eggenberger urns with regard to their asymptotics. In *2005 International Conference on Analysis of Algorithms*, DMTCS Proc. **AD**, pp. 275–285.
[32] Pouyanne, N. (2008) An algebraic approach to Pólya processes. *Ann. Inst. Henri Poincaré Probab. Stat.* **44** 293–323.
[33] Rachev, S. T. and Rüschendorf, L. (1995) Probability metrics and recursive algorithms. *Adv. Appl. Probab.* **27** 770–799.
[34] Rösler, U. (1991) A limit theorem for 'Quicksort'. *RAIRO Inform. Théor. Appl.* **25** 85–100.
[35] Smythe, R. T. (1996) Central limit theorems for urn models. *Stochastic Process. Appl.* **65** 115–137.
[36] Smythe, R. T. and Rosenberger, W. F. (1995) Play-the-winner designs, generalized Pólya urns, and Markov branching processes. In *Adaptive Designs* (N. Flournoy and W. F. Rosenberger, eds), Vol. 25 of *IMS Lecture Notes Monograph Series*, Institute of Mathematical Statistics, pp. 13–22.
[37] Wei, L. J. and Durham, S. (1978). The randomized play-the-winner rule in medical trials. *J. Amer. Statist. Assoc.* **73** 840–843.
[38] Wei, L. J., Smythe, R. T., Lin, D. Y. and Park, T. S. (1990) Statistical inference with data-dependent treatment allocation rules. *J. Amer. Statist. Assoc.* **85** 156–162.
[39] Zolotarev, V. M. (1976) Approximation of the distributions of sums of independent random variables with values in infinite-dimensional spaces (Russian). *Teor. Veroyatnost. i Primenen.* **21** 741–758. Erratum *ibid.* **22** (1977), 901. English translation in *Theory Probab. Appl.* **21** 721–737; *ibid.* **22** 881.
[40] Zolotarev, V. M. (1977) Ideal metrics in the problem of approximating the distributions of sums of independent random variables (Russian). *Teor. Veroyatnost. i Primenen.* **22** 449–465. English translation in *Theory Probab. Appl.* **22** 433–449.