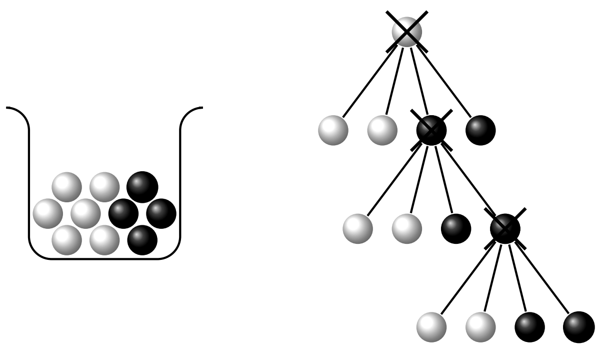
Pólya urns via the contraction method

Dissertation
zur Erlangung des Doktorgrades
der Naturissenschaften

vorgelegt beim Fachbereich Mathematik und Informatik
der Johann Wolfgang Goethe-Universität
in Frankfurt am Main

von
Margarete Carola Knape
aus Lindau (Bodensee)



Frankfurt (2013)
(D30)

vom Fachbereich Mathematik und Informatik der
Johann Wolfgang Goethe-Universität als Dissertation angenommen.

Dekan: Prof. Dr. Thorsten Theobald

Gutachter: Prof. Dr. Ralph Neininger
Prof. Dr. Hosam M. Mahmoud (George Washington University)

Datum der Disputation: 08.11.2013

Acknowledgements

First of all, I wish to express my deepest gratitude to my advisor Ralph Neininger for introducing me to this area of research and his excellent mentoring. I am thankful for his continuous advice and encouragement and his many helpful ideas.

It was a pleasure to work in the lively stochastics group, where throughout the entire time of my work on this topic, I always found willing help for any mathematical, technical, or organizational question.

Last but not least, none of this would have been possible without the love and patience of my family. I would like to express my heart-felt gratitude to my parents for encouraging me throughout my studies. I owe thanks to my brother Georg and my husband Timm for carefully proofreading the thesis. And most of all, I am indebted to my husband Timm and my daughter Ronja for their support, endurance and understanding.

Contents

1 Introduction

1.1 Pólya Urn models

In 1923, George Pólya and Florian Eggenberger published their paper "Über die Statistik verketteter Vorgänge"[11]. They proposed an urn model to describe dependent events and used the monthly number of deaths from smallpox in Switzerland as an example. The urn contains balls in two colors. In each step, a ball is drawn at random from the urn and a fixed number of balls of the same color are added to or removed from the urn. They show that this model fits their data much better than the assumption that the events are independent.

Although similar models had already been discussed before, this model and some extensions thereof are usually called Pólya (or Pólya-Eggenberger) urn models. Such models have been used to model various problems in a wide variety of scientific domains and they have been investigated using various stochastic methods. Interesting accounts of the history of urn models in general and more specifically Pólya urn models can be found for example in the monographs of Johnson and Kotz [18] and Mahmoud [23].

More generally, a Pólya urn model consists of an 'urn' containing a number of 'balls' of different 'colors', and a set of 'rules'. In each step, a ball is drawn at random from the urn, its color is observed and the ball is put back into the urn. Depending on the color of the drawn ball some balls are now added to or removed from the urn according to the given rules. This rules may include some further randomness, for example throwing a coin to determine the color of the added balls, but may not depend on the contents of the urn. If balls are removed, there must always be enough balls in the urn to execute this step.

A classical problem is to identify the asymptotic behavior of the numbers of balls of each color as the number of steps tends to infinity. The literature on this problem, in particular on limit laws for the normalized numbers of balls of each color, is vast. We refer again to the two monographs of Johnson and Kotz [18] and Mahmoud [23] and the references and comments on the literature in the papers of Janson [15], Flajolet et al. [13] and Pouyanne [30].

Several approaches have been used to analyze the asymptotic behavior 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 that the distribution of the composition at time n given time $n-1$ is explicitly accessible.

In this dissertation, an approach based on a "backward" decomposition of the urn process is proposed. We construct an embedding of the evolution of the urn into an associated combinatorial random tree structure growing in discrete time, see chapter 2. 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 color 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 color. To extract the asymptotic behavior from such systems, we develop an approach in the context of the contraction method in chapter 3 and 4.

1.2 Contraction Method

The contraction method is well known in the probabilistic analysis of algorithms. It was introduced by Rösler [32] and first developed systematically in Rachev and Rüschendorf [31]. A rather general framework with numerous applications to the analysis of recursive algorithms and random trees was given by Neininger and Rüschendorf [26]. 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. In this dissertation, a system of such recurrence relations is considered. To the best of our knowledge the method has not yet been used for such systems of recurrence relations, the only exception being Leckey et al. [22] where tries are analyzed under a Markov source model. In the last chapter, an approach more in the spirit of earlier applications of the contraction method is described. We will see that this has some drawbacks compared to the approach developed in chapters 2–4.

1.3 Scope

The aim of this dissertation is not to compete with other techniques with respect to generality under which urn models can be analyzed. Instead, we discuss our approach in a few examples, illustrating the contraction framework in three frequently occurring asymptotic regimes: normal limit laws, non-normal limit laws and regimes with oscillating distributional behavior. 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 be developed into more general theorems when asymptotic expansions of means (respectively means and variances in the normal limit case) are available, cf. the types of expansions of the means in section 3.

We consider an urn with balls in a finite number $m \geq 2$ of different colors, numbered by $1, \ldots, m$. The replacement rules of the urn are encoded by an $m \times m$ replacement matrix $R = (a_{ij})_{1 \leq i,j \leq m}$ 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 color i, it is placed back into the urn together with a_{ij} balls of color j for all $j = 1, \ldots, m$. The steps are iterated.

Throughout this dissertation, we assume that in each step a fixed number $K \geq 2$ of balls are put into the urn.[1] Therefore, the replacement matrix is balanced, i.e.

$$\sum_{j=1}^{m} a_{ij} =: K - 1 \quad \text{for all } i = 1, \ldots, m.$$

For the associated tree process, which lies at the core of our approach, this balance condition implies that asymptotically, the growths of the subtrees can jointly be captured by Dirichlet distributions. This leads to characterizations of the limit distributions in all cases (normal, non-normal and oscillatory limits) by systems,

[1] The notation K is unfortunate since this integer is not random and mainly chosen to have similarity in notation with earlier work on the contraction method.

cf. (3.1)–(3.3) 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. It may be an interesting aspect of the present approach that all three regimes are governed by these quite similar types of systems of distributional fixed-point equations.

1.4 Outline and Notation

In the next chapter, the associated trees are introduced and the embedding of the urn models is described in detail. Furthermore, the systems of distributional recurrences for the numbers of balls of a certain color are derived from the recursive properties of the associated trees.

In chapter 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 chapter 3.1 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 chapter 3.2, 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.

Using these results, we discuss examples of limit laws for Pólya urn schemes within our approach in chapter 4. Furthermore, our convergence proofs are worked out there, again based on the product versions of the minimal L_p and Zolotarev metrics. The contents of chapters 2–4 have been submitted for publication and are available on arXiv.org [20].

In chapter 5 we investigate a variant of our approach, working with one recurrence for random vectors instead of the system of recurrences for (one-dimensional) random variables. This seemed the natural way for the application of the contraction method at first, but led to several problems. After working this out in detail for one of the examples, an extension of the main theorem of Neininger and Rüschendorf [26, Thm 4.1] is presented which might also be useful for other applications of the contraction method.

Notation. By $\xrightarrow{d}$ convergence in distribution is denoted. For the normal distribution on $\mathbb{R}$ with mean $\mu \in \mathbb{R}$ and variance $\sigma^2 \geq 0$, the notation $\mathcal{N}(\mu, \sigma^2)$ is used. In the case $\sigma^2 = 0$, this degenerates to the Dirac measure in μ. Throughout the dissertation, Bachmann-Landau symbols are used in asymptotic statements. By $\log(x)$ for $x > 0$, the natural logarithm of x is denoted. We denote the imaginary unit by i and for any $x = a + \mathrm{i}b \in \mathbb{C}$ we denote its complex conjugate by $\overline{x} \coloneqq a - \mathrm{i}b$.

2 A recursive description of Pólya urns

In this chapter, the embedding of urn processes into associated combinatorial random tree structures growing in discrete time is explained in detail. 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.

2.1 The Pólya urn

For illustration, we first consider an urn model with two colors, black and white, and a deterministic replacement matrix R. In the sequel, an extension of this approach to urns with more than two colors and replacement matrices with random entries is discussed as well. However, the assumption that the sums of the entries in each row are the same is crucial for our method, the reason also being explained below. To be definite, we use the replacement matrix

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

with

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

Hence, 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 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 color are still added to the urn. As initial configurations, we consider both, one black ball or one white ball. We denote by B_n^{b} the number of black balls after n steps when initially starting with one black ball, by B_n^{w} the number of black balls after n steps when initially starting with one white ball. Hence, we have $B_0^{\mathrm{b}} = 1$ and $B_0^{\mathrm{w}} = 0$.

2.2 The associated tree

We encode the urn process by a discrete time evolution of a random tree with nodes colored black or white. This tree is called *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) color. 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 urn, 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 the evolution of an urn and its associated tree see Figure 2.1. Hence, at any time, the balls in the urn are represented by the leaves of the associate tree, where the colors of balls and representing leaves match. Each node of the tree is either a leaf or has K children. We could as well emulate the urn process by only running the evolution of the associated tree as follows: Start with one root node of the color of the initial ball of the urn. At any step, choose one of the leaves of the tree uniformly at random, inspect its color, add K children to the chosen leaf and color 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 distributed 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 with two colors, we consider the two initial compositions of one black 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 $n \geq 1$ steps. It has $n\,(K-1)+1$ leaves,

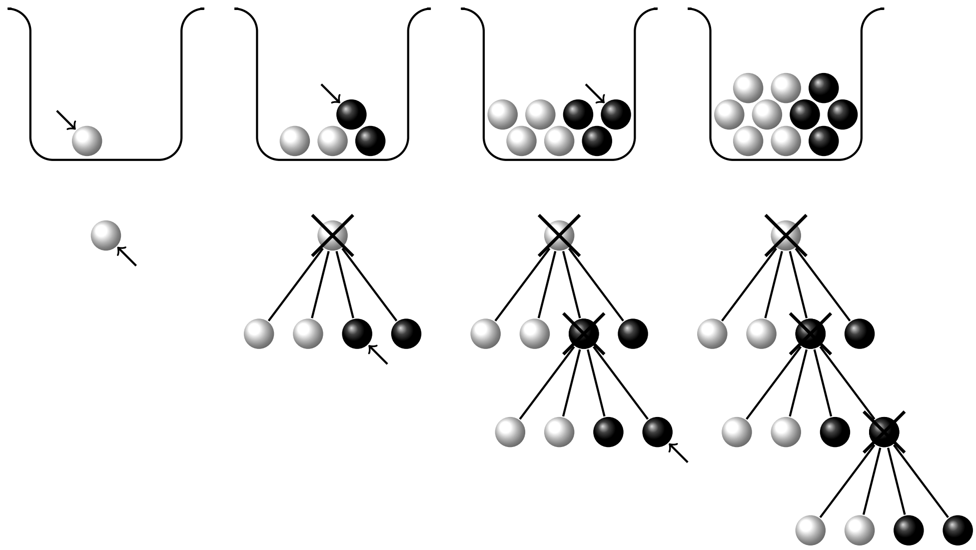


Figure 2.1: A realization of the evolution of the Pólya urn with replacement matrix $\left[\begin{smallmatrix}1 & 2\\ 2 & 1\end{smallmatrix}\right]$ and initially one white ball. The arrows indicate which ball is drawn (resp. node is replaced) in each step. Below each urn its associated tree is shown. Leaf nodes correspond to the balls in the urn, non-leaf nodes (crossed out) do no longer correspond to balls in the urn. However, their color still matters for the recursive decomposition of the associated tree.

and each subtree rooted at a child of the associated tree's root (we call them simply only subtrees) 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 to be $i_r\,(K-1)+1$ 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 they start growing. From the evolution of the b-associated tree, it is clear that, conditioned on the subtrees' numbers of leaves being $i_r\,(K-1)+1$, the subtrees are stochastically independent and the r-th subtree is distributed as an associated tree after i_r steps. Whether it has the distribution of the b- or the w-associated tree depends on the color of the subtree's root node.

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

2.3 Growth of subtrees

In our analysis, the asymptotic growth of the K subtrees of the associated tree is used. We denote by $I^{(n)} = \big(I_1^{(n)}, \dots, I_K^{(n)}\big)$ the vector of the numbers of draws of leaves from each subtree after $n \geq 1$ draws in the full associated tree. In other words, $I_r^{(n)}\,(K-1)+1$ is the number of leaves of the r-th subtree after $n \geq 1$ steps. We have $I^{(1)} = (0, \dots, 0)$, 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 $\mathrm{Dirichlet}\big((K-1)^{-1}, \dots, (K-1)^{-1}\big)$: It is the distribution of a random vector $(D_1, \dots, D_K)$ with $\sum_{r=1}^K D_j = 1$ and such that $(D_1, \dots, D_{K-1})$ has a Lebesgue-density supported by the simplex

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

and given for $x = (x_1, \dots, x_{K-1}) \in \mathcal{S}_K$ by

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

where Γ denotes Euler's Gamma function. In particular, $D_1, \dots, D_K$ are identically distributed, having a $\mathrm{beta}\big((K-1)^{-1}, 1\big)$ 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 behavior of $I^{(n)}$:

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

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

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

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

Proof. The sequence $\left(I_1^{(n)}(K-1)+1,\dots,I_K^{(n)}(K-1)+1\right)_{n\in\mathbb{N}_0}$ 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 $\{1,\dots,K\}$ and we label a leaf j if it belongs to the j-th subtree of the root (any ordering of the subtrees of the root is fine). Hence, all leaves of a subtree of the associated tree's root get the same label, leaves of different subtrees get different labels. The subtree-induced urn now has balls of colors $1,\dots,K$. At any time, the number of balls of each color is identical with the number of leaves with the corresponding label. Hence, the dynamics of the subtree-induced urn are that of a Pólya urn with initially K balls, one of each color. Whenever a ball is drawn, it is placed back into the urn together with $K-1$ balls of the same color. 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 color r. The dynamic of the subtree-induced urn as a K-color Pólya-Eggenberger urn is well-known, cf. Athreya [1, Corollary 1], we have for $n\to\infty$

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

almost surely and in L_p for any $p\geq 1$, where the random vector $(D_1,\dots,D_K)$ has a Dirichlet$\left((K-1)^{-1},\dots,(K-1)^{-1}\right)$ 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 does only enter our subsequent analysis via Lemma 2.1. It seems feasible to apply our approach also 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 $V_1,\dots,V_K\geq 0$ such that $\sum_{r=1}^K V_r=1$ almost surely and $\mathbb{P}[\max_{1\leq r\leq K}V_r<1]>0$. It seems that the contraction argument may even allow that the distribution of V depends on the color of the ball the urn is started with. We leave these issues for future research.

2.4 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 $n \geq 1$, we express as the sum of the numbers of black leaves of its subtrees. As discussed above, conditioned on $I^{(n)} = \big(I_1^{(n)}, \ldots, I_K^{(n)}\big)$, 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 color of their roots. In a b-associated tree, the root has $a+1$ black and $b = K - (a+1)$ white children. Hence, we obtain

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

where $\stackrel{d}{=}$ denotes that left and right hand side have an identical distribution and we have that $(B_k^{\mathrm{b},(1)})_{0\leq k<n}, \ldots, (B_k^{\mathrm{b},(a+1)})_{0\leq k<n}, (B_k^{\mathrm{w},(a+2)})_{0\leq k<n}, \ldots, (B_k^{\mathrm{w},(K)})_{0\leq k<n}$, and $I^{(n)}$ are independent, and for $k = 0, \ldots, n-1$ and the respective values of r, the $B_k^{\mathrm{b},(r)}$ are distributed as B_k^{b} and the $B_k^{\mathrm{w},(r)}$ are distributed as B_k^{w}.

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

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

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 $\big(\mathcal{L}\big(B_n^{\mathrm{b}}\big), \mathcal{L}(B_n^{\mathrm{w}})\big)_{n\geq 0}$.

2.5 Extensions of the model

As mentioned above, our approach can be extended to urn models with more than two different colors. We can also cover the case of a random replacement matrix where for each row, the entries add up to a deterministic and fixed integer $K - 1 \geq 1$.

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

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

The colors (subsequently also called types) are now denoted by numbers $1, \dots, m$ and we focus on the number of balls of color 1 after n steps. When starting with one ball of color j, we denote by $B_n^{[j]}$ the number of color 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

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

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

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

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

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

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

Random entries in the replacement matrix The case of a replacement matrix with random entries such that all rows sum up to a deterministic and fixed $K - 1 \ge 1$ can be covered by an extension of the system (2.5). Instead of a deterministic replacement matrix R, in this case, we are given a distribution on the space of all matrices of the respective size with integer entries such that each row sums up to $K - 1$. For each draw from the urn, a matrix is drawn according to this distribution, independently of everything else. Instead of formulating such an extension explicitly, we discuss an example in section 4.2.

3 Systems of limit equations

We outline how systems of the form (2.5) are used subsequently. Crucial are the expansions of the means

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

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 $\left[\begin{smallmatrix} 4 & 0 \\ 3 & 1 \end{smallmatrix}\right]$, see Kotz et al. [21] for this example or Janson [16] for a comprehensive account of urns with triangular replacement matrix.

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

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

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 of **type (a)**. This suggests that the variances are of the order $n^{2\lambda}$ and a proper scaling is

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

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 chapter 4), we obtain the system of fixed-point equations

$$X^{[j]} \stackrel{d}{=} \sum_{i=1}^{m} \sum_{r \in J_{ij}} D_r^\lambda X^{[i],(r)} + b^{[j]}, \quad j = 1, \dots, m, \tag{3.1}$$

where all $X^{[i],(r)}$ and $(D_1, \dots, D_K)$ are independent, each $X^{[i],(r)}$ is distributed as $X^{[i]}$, $(D_1, \dots, D_K)$ is distributed as in Lemma 2.1, and each $b^{[j]}$ is a function of

$(D_1, \ldots, D_K)$. It turns out that such a system, when restricted to centered $X^{[j]}$ with finite second moments, has a unique solution on the level of distributions (Theorem 3.1). This identifies the weak limits of the $X_n^{[j]}$. Examples can be found in sections 4.1 and 4.2. One can as well obtain the same system (3.1) with $b^{[j]} = 0$ for all j by only centering the $B_n^{[j]}$ by $c_\mu n$ instead of the exact mean. In this case, system (3.1) has to be solved subject to finite second moments and appropriate means.

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

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

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

$$X_n^{[j]} := \frac{B_n^{[j]} - \mu_n^{[j]}}{\sqrt{\mathrm{Var}\big(B_n^{[j]}\big)}}, \quad n \geq 1, \quad j = 1, \ldots, m$$

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

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

with conditions as in (3.1). Under appropriate assumptions on moments, we find that the only solution is all $X^{[j]}$ being standard normally distributed (Theorem 3.2). This leads to asymptotic normality of the $X_n^{[j]}$. Examples are given in sections 4.1 and 4.2. The case

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

leads to the same system of fixed-point equations (3.2). However, here the variances typically are of order $n \log^\delta(n)$ with a positive δ.

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

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

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

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

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

$$X^{[j]} \stackrel{d}{=} \sum_{i=1}^{m} \sum_{r \in J_{ij}} D_r^\omega X^{[i],(r)}, \quad j = 1, \ldots, m, \tag{3.3}$$

with conditions as in (3.1) and $\omega := \lambda + \mathrm{i}\mu$. Under appropriate moment assumptions, this has a unique solution within distributions on $\mathbb{C}$ (Theorem 3.3). An example of a corresponding distributional approximation is given in section 4.3.

Note that the approach of embedding urn models into continuous time multitype branching processes, see [2, 15], also leads to characterizations of the non-normal limits similar to (3.1) and (3.3). However, the form of the fixed-point equations is different, see the system in equation (3.5) in Janson [15]. Properties of such fixed-points have been studied in Chauvin et al. [9, 7, 8].

3.1 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 to first characterize limit distributions of urn models (section 3.2) and then prove convergence in distribution of the scaled numbers of balls of a certain color (section 4).

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

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

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

$$\left(\mathcal{M}_s^{\mathbb{R}}\right)^{\times d} := \underbrace{\mathcal{M}_s^{\mathbb{R}} \times \cdots \times \mathcal{M}_s^{\mathbb{R}}}_{d \text{ times}},$$

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

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

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

To cover the different behavior 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 has been introduced and studied in [37, 38]. The contraction method based on the Zolotarev metric was systematically developed in [26] and, for issues that go beyond what is needed in this paper, in [19] and [28]. 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

$$\zeta_s(X,Y) := \zeta_s\big(\mathcal{L}(X)\,,\mathcal{L}(Y)\big) := \sup_{f\in\mathcal{F}_s}\Big|\mathbb{E}\big[f(X)-f(Y)\big]\Big|\,, \tag{3.4}$$

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

$$\mathcal{F}_s := \left\{ f \in C^m(\mathbb{R},\mathbb{R}) \;\middle|\; \Big|f^{(m)}(x) - f^{(m)}(y)\Big| \le |x-y|^\alpha \right\}, \tag{3.5}$$

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

We have $\zeta_s(X,Y) < \infty$, if all moments of orders $1,\dots,m$ of X and Y are equal and if the s-th absolute moments of X and Y are finite. Since later on the cases $1 < s \le 3$ are used, we have two basic cases: First, for $1 < s \le 2$, 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 $2 < s \le 3$, we have $\zeta_s(X,Y) < \infty$ for $\mathcal{L}(X)$, $\mathcal{L}(Y) \in \mathcal{M}_s^{\mathbb{R}}\big(\mu,\sigma^2\big)$ for any $\mu \in \mathbb{R}$ and $\sigma \ge 0$. Moreover, the pairs $\big(\mathcal{M}_s^{\mathbb{R}}(\mu)\,,\zeta_s\big)$ for $1 < s \le 2$ and $\big(\mathcal{M}_s^{\mathbb{R}}\big(\mu,\sigma^2\big)\,,\zeta_s\big)$ for $2 < s \le 3$ are complete metric spaces; for the completeness see [10, Theorem 5.1].

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

$$\zeta_s(X+Z, Y+Z) \le \zeta_s(X,Y)\,, \quad \zeta_s(cX,cY) = c^s\zeta_s(X,Y)\,, \tag{3.6}$$

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

$$\zeta_s\left(\sum_{i=1}^n X_i,\, \sum_{i=1}^n Y_i\right) \le \sum_{i=1}^n \zeta_s(X_i,\, Y_i)\,. \tag{3.7}$$

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

$$\zeta_s^{\vee}\big((\nu_1,\dots,\nu_d),(\mu_1,\dots,\mu_d)\big) := \max_{1\le j\le d} \zeta_s(\nu_j,\mu_j), \tag{3.8}$$

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

Minimal $\mathbf{L_p}$-metric $\boldsymbol{\ell_p}$. First, for probability metrics on the real line, the minimal L_p-metric ℓ_p, $1 \le p < \infty$ is defined by

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

where $\|V - W\|_p := \left(\mathbb{E}\left[|V - W|^p\right]\right)^{1/p}$ is the usual L_p-norm. The spaces $\left(\mathcal{M}_p, \ell_p\right)$ and $\left(\mathcal{M}_p(\mu), \ell_p\right)$ for $1 \le p < \infty$ are complete metric spaces, see [6]. The infimum in the definition of ℓ_p is in fact 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 do exist for all $\nu, \varrho \in \mathcal{M}_1^{\mathbb{R}}$. We use the notation $\ell_p(X, Y) := \ell_p(\mathcal{L}(X), \mathcal{L}(Y))$ for random variables X and Y. Subsequently also the following inequality between the ℓ_p and ζ_s metrics is used (see [10, Lemma 5.7]):

$$\zeta_s(X, Y) \le \left(\left(\mathbb{E}\left[|X|^s\right]\right)^{1-1/s} + \left(\mathbb{E}\left[|Y|^s\right]\right)^{1-1/s} \right) \ell_s(X, Y), \quad 1 < s \le 3, \tag{3.9}$$

where, for $1 < s \le 2$, we need $\mathcal{L}(X), \mathcal{L}(Y) \in \mathcal{M}_s^{\mathbb{R}}(\mu)$ for some $\mu \in \mathbb{R}$ and, for $2 < s \le 3$, we need $\mathcal{L}(X), \mathcal{L}(Y) \in \mathcal{M}_s^{\mathbb{R}}(\mu, \sigma^2)$ for some $\mu \in \mathbb{R}$ and $\sigma \ge 0$.

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

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

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

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

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

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 $\left(\mathcal{M}_2^{\mathbb{C}}(0)\right)^{\times d}$, we use

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

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

Preview on the use of spaces and metrics. The guidance on which space and metric to use in which asymptotic regime of Pólya urns is as follows. We come back 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, however 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 behavior of the distributions. Such oscillatory behavior can be captured almost surely, we discuss a (weak) description for distributions.

The cases of **type (a)** can be dealt with on the space $\big(\mathcal{M}_2^{\mathbb{R}}(\mu)\big)^{\times d}$ with appropriate $\mu \in \mathbb{R}$ and $d \in \mathbb{N}$, where, by centering, one can always achieve the choice $\mu = 0$. One can either use the metric $\zeta_2^{\vee}$ 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 do not see any advantage of $\ell_2^{\vee}$ here.

The cases of **type (b)** can be dealt with on the space $\big(\mathcal{M}_s^{\mathbb{R}}(\mu, \sigma^2)\big)^{\times d}$ with $2 < s \leq 3$ 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 do 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 $\big(\mathcal{M}_2^{\mathbb{C}}(\gamma)\big)^{\times d}$ with appropriate $\gamma \in \mathbb{R}$ and $d \in \mathbb{N}$. The metric of choice for the 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 $\gamma_1, \ldots, \gamma_d \in \mathbb{C}$ to be able to work with a more natural scaling of the random variables, the metric still being $\ell_2^{\vee}$.

3.2 Associated fixed-point equations

We fix $d, d' \in \mathbb{N}$, a $d \times d'$ matrix $\big(A_{ir}\big)$ 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 $\big(\pi(ir)\big)$ with all entries $\pi(ir) \in \{1, \dots, d\}$. First, we consider the case where all A_{ir} and all b_i are real. We associate a map

$$\begin{aligned} T : \left(\mathcal{M}^{\mathbb{R}}\right)^{\times d} &\to \left(\mathcal{M}^{\mathbb{R}}\right)^{\times d} \\ (\mu_1, \dots, \mu_d) &\mapsto \big(T_1(\mu_1, \dots, \mu_d), \dots, T_d(\mu_1, \dots, \mu_d)\big) \qquad (3.10) \\ T_i(\mu_1, \dots, \mu_d) &:= \mathcal{L}\Bigg(\sum_{r=1}^{d'} A_{ir} Z_{ir} + b_i\Bigg) \qquad (3.11) \end{aligned}$$

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

$$\begin{aligned} T' : \left(\mathcal{M}^{\mathbb{C}}\right)^{\times d} &\to \left(\mathcal{M}^{\mathbb{C}}\right)^{\times d} \qquad (3.12) \\ (\mu_1, \dots, \mu_d) &\mapsto \big(T'_1(\mu_1, \dots, \mu_d), \dots, T'_d(\mu_1, \dots, \mu_d)\big) \end{aligned}$$

with $T'_i(\mu_1, \dots, \mu_d)$ defined as for T_i in (3.11).

For the three regimes discussed in the preview within section 3.1, we use the following three theorems (Theorem 3.1 for type (a), Theorem 3.2 for type (b), and Theorem 3.3 for type (c)) on existence of fixed-points of T and T'.

Theorem 3.1. *Assume that in the definition of T in (3.10) and (3.11) the A_{ir} and b_i are square integrable real random variables with $\mathbb{E}[b_i] = 0$ for all $1 \le i \le d$ and $1 \le r \le d'$ and*

$$\max_{1 \le i \le d} \sum_{r=1}^{d'} \mathbb{E}\left[A_{ir}^2\right] < 1. \qquad (3.13)$$

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

Theorem 3.2. *Assume that in the definition of T in (3.10) and (3.11) for some $\varepsilon > 0$ the A_{ir} are $L_{2+\varepsilon}$-integrable real random variables and $b_i = 0$ for all $1 \le i \le d$ and $1 \le r \le d'$, and that*

$$\sum_{r=1}^{d'} A_{ir}^2 = 1 \quad \text{for all } i = 1, \ldots, d, \tag{3.14}$$

and

$$\min_{1 \le i \le d} \mathbb{P}\left(\max_{1 \le r \le d'} |A_{ir}| < 1 \right) > 0. \tag{3.15}$$

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

Theorem 3.3. *Assume that in the definition of T' in (3.12), the A_{ir} and b_i are square integrable complex random variables for all $1 \le i \le d$ and $1 \le r \le d'$, and that, for $\gamma_1, \ldots, \gamma_d \in \mathbb{C}$, we have*

$$\mathbb{E}[b_i] + \sum_{r=1}^{d'} \gamma_{\pi(ir)} \mathbb{E}[A_{ir}] = \gamma_i, \quad i = 1, \ldots, d. \tag{3.16}$$

If moreover

$$\max_{1 \le i \le d} \sum_{r=1}^{d'} \mathbb{E}\left[|A_{ir}|^2 \right] < 1, \tag{3.17}$$

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 3.1 was used in the proof of [15, Thm. 3.9 (iii)] with a similar proof technique as in our proof of Theorem 3.3.

The rest of this section contains the proofs of Theorems 3.1–3.3.

Proof (Theorem 3.1). First note that for $(\mu_1, \ldots, \mu_d) \in \left(\mathcal{M}_2^{\mathbb{R}}(0)\right)^{\times d}$, by independence in definition (3.11) and $\mathbb{E}[b_i] = 0$, we have $T_i(\mu_1, \ldots, \mu_d) \in \mathcal{M}_2^{\mathbb{R}}(0)$ for $i = 1, \ldots, d$. Hence, the restriction of T to $\left(\mathcal{M}_2^{\mathbb{R}}(0)\right)^{\times d}$ maps into $\left(\mathcal{M}_2^{\mathbb{R}}(0)\right)^{\times d}$.

Next, we show that the restriction of T to $\left(\mathcal{M}_2^{\mathbb{R}}(0)\right)^{\times d}$ is a (strict) contraction with respect to the metric $\zeta_2^{\vee}$: For $(\mu_1, \ldots, \mu_d), (\nu_1, \ldots, \nu_d) \in \left(\mathcal{M}_2^{\mathbb{R}}(0)\right)^{\times d}$, we first fix $i \in \{1, \ldots, d\}$. Let $Z_{i1}, \ldots, Z_{id'}$ and $Z'_{i1}, \ldots, Z'_{id'}$ be real random variables such that

Z_{ir} is distributed as $\mu_{\pi(ir)}$ and Z'_{ir} is distributed as $\nu_{\pi(ir)}$. Moreover, assume that both families $\{(A_{i1},\dots,A_{id'},b_i),\, Z_{i1},\dots,Z_{id'}\}$ and $\{(A_{i1},\dots,A_{id'},b_i),\, Z'_{i1},\dots,Z'_{id'}\}$ are independent. Then we have

$$\begin{aligned} T_i(\mu_1,\dots,\mu_d) &= \mathcal{L}\Big(\sum_{r=1}^{d'} A_{ir}Z_{ir} + b_i\Big), \\ T_i(\nu_1,\dots,\nu_d) &= \mathcal{L}\Big(\sum_{r=1}^{d'} A_{ir}Z'_{ir} + b_i\Big). \end{aligned} \tag{3.18}$$

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

$$\begin{aligned} &\zeta_2\big(T_i(\mu_1,\dots,\mu_d), T_i(\nu_1,\dots,\nu_d)\big) \\ &\quad= \sup_{f\in\mathcal{F}_2}\left|\int f\Big(\sum_{r=1}^{d'}\alpha_r Z_{ir}+\beta\Big) - f\Big(\sum_{r=1}^{d'}\alpha_r Z'_{ir}+\beta\Big)\,d\Upsilon(\alpha_1,\dots,\alpha_{d'},\beta)\right| \\ &\quad\le \int \sup_{f\in\mathcal{F}_2}\left|f\Big(\sum_{r=1}^{d'}\alpha_r Z_{ir}+\beta\Big) - f\Big(\sum_{r=1}^{d'}\alpha_r Z'_{ir}+\beta\Big)\right|\,d\Upsilon(\alpha_1,\dots,\alpha_{d'},\beta) \\ &\quad= \int \zeta_2\Big(\sum_{r=1}^{d'}\alpha_r Z_{ir}+\beta,\ \sum_{r=1}^{d'}\alpha_r Z'_{ir}+\beta\Big)\,d\Upsilon(\alpha_1,\dots,\alpha_{d'},\beta) \end{aligned} \tag{3.19}$$

Since ζ_2 is $(2,+)$ ideal, we obtain from (3.6) that

$$\zeta_2\Big(\sum_{r=1}^{d'}\alpha_r Z_{ir}+\beta,\ \sum_{r=1}^{d'}\alpha_r Z'_{ir}+\beta\Big) \le \sum_{r=1}^{d'}\alpha_r^2\,\zeta_2\big(Z_{ir},Z'_{ir}\big)\,. \tag{3.20}$$

Hence, we can further estimate

$$\begin{aligned} &\zeta_2\big(T_i(\mu_1,\dots,\mu_d)\,,T_i(\nu_1,\dots,\nu_d)\big) \\ &\qquad\le \int \sum_{r=1}^{d'}\alpha_r^2\,\zeta_2\big(Z_{ir},Z'_{ir}\big)\ d\Upsilon(\alpha_1,\dots,\alpha_{d'},\beta) \\ &\qquad= \int \sum_{r=1}^{d'}\alpha_r^2\,\zeta_2\big(\mu_{\pi(ir)},\nu_{\pi(ir)}\big)\ d\Upsilon(\alpha_1,\dots,\alpha_{d'},\beta) \\ &\qquad\le \Big(\sum_{r=1}^{d'}\mathbb{E}\big[A_{ir}^2\big]\Big)\,\zeta_2^{\vee}\big((\mu_1,\dots,\mu_d),(\nu_1,\dots,\nu_d)\big). \end{aligned} \tag{3.21}$$

Now, taking the maximum over i yields

$$\zeta_2^{\vee}\big(T(\mu_1,\dots,\mu_d),\, T(\nu_1,\dots,\nu_d)\big) \leq \left(\max_{1\leq i\leq d}\sum_{r=1}^{d'}\mathbb{E}\big[A_{ir}^2\big]\right)\zeta_2^{\vee}\big((\mu_1,\dots,\mu_d),\,(\nu_1,\dots,\nu_d)\big). \quad (3.22)$$

Hence, condition (3.13) implies that the restriction of T to $\big(\mathcal{M}_2^{\mathbb{R}}(0)\big)^{\times d}$ is a contraction. Since the metric $\zeta_2^{\vee}$ is complete, Banach's fixed-point theorem implies the assertion. □

Proof (Theorem 3.2). This proof is similar to the previous proof of Theorem 3.1. Let $\varepsilon > 0$ be as in Theorem 3.2 and $\sigma > 0$ be arbitrary. First note that for $(\mu_1,\dots,\mu_d) \in \big(\mathcal{M}_{2+\varepsilon}^{\mathbb{R}}\big(0,\sigma^2\big)\big)^{\times d}$, by independence in definition (3.11), condition (3.14), and $b_i = 0$, we have $T_i(\mu_1,\dots,\mu_d) \in \mathcal{M}_{2+\varepsilon}^{\mathbb{R}}\big(0,\sigma^2\big)$ for $i = 1,\dots,d$. Hence, the restriction of T to $\big(\mathcal{M}_{2+\varepsilon}^{\mathbb{R}}\big(0,\sigma^2\big)\big)^{\times d}$ maps into $\big(\mathcal{M}_{2+\varepsilon}^{\mathbb{R}}\big(0,\sigma^2\big)\big)^{\times d}$.

We set $s := (2+\varepsilon)\wedge 3$. For $(\mu_1,\dots,\mu_d),(\nu_1,\dots,\nu_d) \in \big(\mathcal{M}_{2+\varepsilon}^{\mathbb{R}}\big(0,\sigma^2\big)\big)^{\times d}$, we choose $Z_{i1},\dots,Z_{id'}$ and $Z'_{i1},\dots,Z'_{id'}$ as in the proof of Theorem 3.1 such that we have (3.18). Note that with our choice of s, we have $\zeta_s\big(T_i(\mu_1,\dots,\mu_d)\,,T_i(\nu_1,\dots,\nu_d)\big) < \infty$. With an estimate analogous to (3.19)–(3.22), using now that ζ_s is $(s,+)$ ideal, we obtain

$$\zeta_s^{\vee}\big(T(\mu_1,\dots,\mu_d),\, T(\nu_1,\dots,\nu_d)\big) \leq \left(\max_{1\leq i\leq d}\sum_{r=1}^{d'}\mathbb{E}\big[|A_{ir}|^s\big]\right)\zeta_s^{\vee}\big((\mu_1,\dots,\mu_d),(\nu_1,\dots,\nu_d)\big). \quad (3.23)$$

Note that $s > 2$ and the conditions (3.14) and (3.15) imply that $\sum_{r=1}^{d'}\mathbb{E}[|A_{ir}|^s] < 1$ for all $i = 1,\dots,d$. Hence, the restriction of T to $\big(\mathcal{M}_{2+\varepsilon}^{\mathbb{R}}\big(0,\sigma^2\big)\big)^{\times d}$ is a contraction and the completeness of $\zeta_s^{\vee}$ implies the existence of a unique fixed-point. Using the convolution property $\mathcal{N}\big(0,\sigma_1^2\big) * \mathcal{N}\big(0,\sigma_2^2\big) = \mathcal{N}\big(0,\sigma_1^2+\sigma_2^2\big)$ for $\sigma_1,\sigma_2 \geq 0$, one can directly check that $\big(\mathcal{N}\big(0,\sigma^2\big),\dots,\mathcal{N}\big(0,\sigma^2\big)\big)$ is a fixed-point of T in the space $\big(\mathcal{M}_{2+\varepsilon}^{\mathbb{R}}\big(0,\sigma^2\big)\big)^{\times d}$. □

Proof (Theorem 3.3). Let $\gamma_1,\dots,\gamma_d$ be as stated in Theorem 3.3 and abbreviate $\mathcal{P} := \mathcal{M}_2^{\mathbb{C}}(\gamma_1)\times\dots\times\mathcal{M}_2^{\mathbb{C}}(\gamma_d)$. 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

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

by condition (3.16). 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 $\mu_{\pi(ir)}$ and $\nu_{\pi(ir)}$ 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

$$\begin{aligned} T_i'(\mu_1,\dots,\mu_d) &= \mathcal{L}\Bigg(\sum_{r=1}^{d'} A_{ir} Z_{ir} + b_i\Bigg), \\ T_i'(\nu_1,\dots,\nu_d) &= \mathcal{L}\Bigg(\sum_{r=1}^{d'} A_{ir} Z_{ir}' + b_i\Bigg). \end{aligned} \tag{3.24}$$

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

$$\begin{aligned} &\ell_2^2\big(T_i'(\mu_1,\dots,\mu_d)\,,T_i'(\nu_1,\dots,\nu_d)\big) \\ &\qquad \le \mathbb{E}\left[\left|\sum_{r=1}^{d'} A_{ir}\left(Z_{ir} - Z_{ir}'\right)\right|^2\right] \\ &\qquad = \mathbb{E}\left[\sum_{r=1}^{d'} |A_{ir}|^2 \left|Z_{ir} - Z_{ir}'\right|^2\right] + \mathbb{E}\left[\sum_{r\neq t} A_{ir}\left(Z_{ir} - Z_{ir}'\right)\overline{A_{it}\left(Z_{it} - Z_{it}'\right)}\right] \\ &\qquad = \sum_{r=1}^{d'} \mathbb{E}\Big[|A_{ir}|^2\Big]\,\ell_2^2\big(\mu_{\pi(ir)},\nu_{\pi(ir)}\big) \qquad\qquad (3.25) \\ &\qquad \le \left(\sum_{r=1}^{d'} \mathbb{E}\left[|A_{ir}|^2\right]\right)\Big(\ell_2^{\vee}\big((\mu_1,\dots,\mu_d),(\nu_1,\dots,\nu_d)\big)\Big)^2. \end{aligned}$$

For equality (3.25) we firstly use that $Z_{ir} - Z_{ir}'$ and $Z_{it} - Z_{it}'$ are independent and centered factors, implying that the expectation of the sum over all $r \neq t$ is 0. Furthermore, (Z_{ir}, Z_{ir}') are optimal couplings of $\big(\mu_{\pi(ir)},\nu_{\pi(ir)}\big)$ which in turn ensures that $\mathbb{E}\big[|Z_{ir} - Z_{ir}'|^2\big] = \ell_2^2(\mu_{\pi(ir)},\nu_{\pi(ir)})$.

Now, taking the maximum over i yields

$$\ell_2^\vee\big(T'(\mu_1,\ldots,\mu_d),T'(\nu_1,\ldots,\nu_d)\big) \leq \left(\max_{1\leq i\leq d}\sum_{r=1}^{d'}\mathbb{E}\Big[|A_{ir}|^2\Big]\right)^{1/2}\ell_2^\vee\big((\mu_1,\ldots,\mu_d),(\nu_1,\ldots,\nu_d)\big). \quad (3.26)$$

Hence, condition (3.17) 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. □

4 Convergence and examples

In this chapter, a couple of concrete Pólya urns are considered and convergence of the normalized numbers of balls of a color is shown with respect to the product metrics defined in chapter 3.1. The proofs are generic in order that they can easily be applied to other urns of the types (a)–(c) in chapter 3. We always consider limit laws for the initial compositions of the urn with one ball of (arbitrary) color. 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. The details are left to the reader.

4.1 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 behavior occurs for these urns when $a - c \leq (a+b)/2$ (type (b) in chapter 3.1), whereas $a - c > (a+b)/2$ leads to limit laws with non-normal limit distributions (type (a) in chapter 3.1). In this chapter we show how to derive these results by our contraction approach. With B_n^{b} and B_n^{w} as in the beginning of chapter 2 we denote expectations by $\mu_{\mathrm{b}}(n)$ and $\mu_{\mathrm{w}}(n)$. These values can be derived exactly, see [3],

$$\mu_{\mathrm{b}}(n) = \frac{c\,(a+b)}{b+c} n + \frac{b\,\Gamma\left(\frac{1}{a+b}\right)}{(b+c)\,\Gamma\left(\frac{1+a-c}{a+b}\right)} \frac{\Gamma\left(n + \frac{1+a-c}{a+b}\right)}{\Gamma\left(n + \frac{1}{a+b}\right)} + \frac{c}{b+c}, \tag{4.1}$$

$$\mu_{\mathrm{w}}(n) = \frac{c\,(a+b)}{b+c} n - \frac{c\,\Gamma\left(\frac{1}{a+b}\right)}{(b+c)\,\Gamma\left(\frac{1+a-c}{a+b}\right)} \frac{\Gamma\left(n + \frac{1+a-c}{a+b}\right)}{\Gamma\left(n + \frac{1}{a+b}\right)} + \frac{c}{b+c}. \tag{4.2}$$

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 $n \to \infty$,

$$\mu_{\mathrm{b}}(n) = c_{\mathrm{b}} n + d_{\mathrm{b}} n^{\lambda} + o(n^{\lambda}), \quad \mu_{\mathrm{w}}(n) = c_{\mathrm{w}} n + d_{\mathrm{w}} n^{\lambda} + o(n^{\lambda}) \tag{4.3}$$

with

$$c_{\mathrm{b}} = c_{\mathrm{w}} = \frac{c\,(a+b)}{b+c}, \quad d_{\mathrm{b}} = \frac{b\,\Gamma\left(\frac{1}{a+b}\right)}{(b+c)\,\Gamma\left(\frac{1+a-c}{a+b}\right)}, \quad d_{\mathrm{w}} = -\frac{c\,\Gamma\left(\frac{1}{a+b}\right)}{(b+c)\,\Gamma\left(\frac{1+a-c}{a+b}\right)}. \tag{4.4}$$

We use the normalizations $X_0 := Y_0 := 0$ and

$$X_n := \frac{B_n^{\mathrm{b}} - \mu_{\mathrm{b}}(n)}{n^{\lambda}}, \qquad Y_n := \frac{B_n^{\mathrm{w}} - \mu_{\mathrm{w}}(n)}{n^{\lambda}}, \qquad n \geq 1. \tag{4.5}$$

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

$$X_n \stackrel{d}{=} \sum_{r=1}^{a+1} \left(\frac{I_r^{(n)}}{n}\right)^{\lambda} X^{(r)}_{I_r^{(n)}} + \sum_{r=a+2}^{K} \left(\frac{I_r^{(n)}}{n}\right)^{\lambda} Y^{(r)}_{I_r^{(n)}} + b_{\mathrm{b}}(n)\,, \tag{4.6}$$

$$Y_n \stackrel{d}{=} \sum_{r=1}^{c} \left(\frac{I_r^{(n)}}{n}\right)^{\lambda} X^{(r)}_{I_r^{(n)}} + \sum_{r=c+1}^{K} \left(\frac{I_r^{(n)}}{n}\right)^{\lambda} Y^{(r)}_{I_r^{(n)}} + b_{\mathrm{w}}(n)\,, \tag{4.7}$$

where

$$b_{\mathrm{b}}(n) = d_{\mathrm{b}} \left(-1 + \sum_{r=1}^{a+1} \left(\frac{I_r^{(n)}}{n}\right)^{\lambda}\right) + d_{\mathrm{w}} \sum_{r=a+2}^{K} \left(\frac{I_r^{(n)}}{n}\right)^{\lambda} + \mathrm{o}(1)\,, \tag{4.8}$$

$$b_{\mathrm{w}}(n) = d_{\mathrm{b}} \sum_{r=1}^{c} \left(\frac{I_r^{(n)}}{n}\right)^{\lambda} + d_{\mathrm{w}} \left(-1 + \sum_{r=c+1}^{K} \left(\frac{I_r^{(n)}}{n}\right)^{\lambda}\right) + \mathrm{o}(1)\,, \tag{4.9}$$

with conditions on independence and identical distributions analogously to (2.2) and (2.3).

In view of Lemma 2.1, this suggests for limits X and Y of X_n and Y_n respectively,

$$X \stackrel{d}{=} \sum_{r=1}^{a+1} D_r^{\lambda} X^{(r)} + \sum_{r=a+2}^{K} D_r^{\lambda} Y^{(r)} + b_{\mathrm{b}}, \tag{4.10}$$

$$Y \stackrel{d}{=} \sum_{r=1}^{c} D_r^{\lambda} X^{(r)} + \sum_{r=c+1}^{K} D_r^{\lambda} Y^{(r)} + b_{\mathrm{w}}, \tag{4.11}$$

with

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

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

where $(D_1, \dots, D_K)$, $X^{(1)}, \dots, X^{(K)}$, $Y^{(1)}, \dots, Y^{(K)}$ are independent, and the random variables $X^{(r)}$ are distributed as X, the r.v. $Y^{(r)}$ are distributed as Y, and $(D_1, \dots, D_K)$ is as in Lemma 2.1. Note that the moments $\mathbb{E}\big[D_r^{\lambda}\big]$ and the form of d_{b} and d_{w} given in (4.4) imply $\mathbb{E}[b_{\mathrm{b}}] = \mathbb{E}[b_{\mathrm{w}}] = 0$. From $\lambda > 1/2$ and $\sum_{r=1}^{K} D_r = 1$, we obtain

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

Hence, Theorem 3.1 can be applied to the map associated with the system (4.10)–(4.11) and implies that there exists a unique solution $\big(\mathcal{L}(\Lambda_{\mathrm{b}}), \mathcal{L}(\Lambda_{\mathrm{w}})\big)$ in the space $\mathcal{M}_2^{\mathbb{R}}(0) \times \mathcal{M}_2^{\mathbb{R}}(0)$. The following convergence proof resembles ideas from Neininger and Rüschendorf [26].

Theorem 4.1. *Consider a Pólya urn with replacement matrix* (2.1) *for which* $a - c > (a+b)/2$ *and* $bc > 0$. *Furthermore, let* X_n *and* Y_n *be the normalized numbers of black balls as in (4.5). We denote by* $(\mathcal{L}(\Lambda_{\mathrm{b}}), \mathcal{L}(\Lambda_{\mathrm{w}}))$ *the solution of (4.10)–(4.11) which is unique in* $\mathcal{M}_2^{\mathbb{R}}(0) \times \mathcal{M}_2^{\mathbb{R}}(0)$. *Then, as* $n \to \infty$,

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

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

$$X_n \xrightarrow{d} \Lambda_{\mathrm{b}}, \quad Y_n \xrightarrow{d} \Lambda_{\mathrm{w}}. \tag{4.12}$$

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

$$Q_n^{\mathrm{b}} := \sum_{r=1}^{a+1} \left(\frac{I_r^{(n)}}{n}\right)^{\lambda} \Lambda_{\mathrm{b}}^{(r)} + \sum_{r=a+2}^{K} \left(\frac{I_r^{(n)}}{n}\right)^{\lambda} \Lambda_{\mathrm{w}}^{(r)} + b_{\mathrm{b}}(n)\,, \tag{4.13}$$

$$Q_n^{\mathrm{w}} := \sum_{r=1}^{c} \left(\frac{I_r^{(n)}}{n}\right)^{\lambda} \Lambda_{\mathrm{b}}^{(r)} + \sum_{r=c+1}^{K} \left(\frac{I_r^{(n)}}{n}\right)^{\lambda} \Lambda_{\mathrm{w}}^{(r)} + b_{\mathrm{w}}(n)\,, \tag{4.14}$$

with $b_{\mathrm{b}}(n)$ and $b_{\mathrm{w}}(n)$ as in (4.8) and the $\Lambda_{\mathrm{b}}^{(r)}$, $\Lambda_{\mathrm{w}}^{(r)}$ and $I^{(n)}$ being independent. All $\Lambda_{\mathrm{b}}^{(r)}$ are distributed as Λ_{b} and all $\Lambda_{\mathrm{w}}^{(r)}$ are distributed as Λ_{w} for the respective values of r. Note that Q_n^{b} and Q_n^{w} are centered with finite second moment since $\mathcal{L}(\Lambda_{\mathrm{b}}), \mathcal{L}(\Lambda_{\mathrm{b}}) \in \mathcal{M}_2^{\mathbb{R}}(0)$. Hence, ζ_2 distances between $X_n, Y_n, Q_n^{\mathrm{b}}, Q_n^{\mathrm{w}}, \Lambda_{\mathrm{b}}$, and Λ_{w} are finite. To bound

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

we look at the distances

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

We start with the estimate

$$\zeta_2(X_n, \Lambda_{\mathrm{b}}) \leq \zeta_2\Big(X_n, Q_n^{\mathrm{b}}\Big) + \zeta_2\Big(Q_n^{\mathrm{b}}, \Lambda_{\mathrm{b}}\Big). \tag{4.15}$$

We first show for the second summand in (4.15) that $\zeta_2\big(Q_n^{\mathrm{b}}, \Lambda_{\mathrm{b}}\big) \to 0$, as $n \to \infty$: With inequality (3.9), we have

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

Moreover, $\|\Lambda_{\mathrm{b}}\|_2$ is finite since $\mathcal{L}(\Lambda_{\mathrm{b}}) \in \mathcal{M}_2^{\mathbb{R}}$. By definition of Q_n^{b} and using that $\Big|I_r^{(n)}/n\Big| \leq 1$, we get that $\big\|Q_n^{\mathrm{b}}\big\|_2$ is uniformly bounded in n. Hence, it is sufficient to show $\ell_2(Q_n^{\mathrm{b}}, \Lambda_{\mathrm{b}}) \to 0$. The independence properties in (4.13) and (4.10) imply that

$$\ell_2\Big(Q_n^{\mathrm{b}}, \Lambda_{\mathrm{b}}\Big) \leq \sum_{r=1}^{a+1} \left\| \left(\frac{I_r^{(n)}}{n}\right)^{\lambda} - D_r^{\lambda} \right\|_2 \Big\|\Lambda_{\mathrm{b}}^{(r)}\Big\|_2 + \sum_{r=a+2}^{K} \left\| \left(\frac{I_r^{(n)}}{n}\right)^{\lambda} - D_r^{\lambda} \right\|_2 \Big\|\Lambda_{\mathrm{w}}^{(r)}\Big\|_2 + \big\|b_{\mathrm{b}}(n) - b_{\mathrm{b}}\big\|_2. \tag{4.16}$$

Lemma 2.1 implies that $\left\|\left(I_r^{(n)}/n\right)^\lambda - D_r^\lambda\right\|_2 \to 0$ as $n \to \infty$, which in turn implies $\left\|b_\mathrm{b}(n) - b_\mathrm{b}\right\|_2 \to 0$. Hence, we obtain $\ell_2(Q_n^\mathrm{b}, \Lambda_\mathrm{b}) \to 0$ and $\zeta_2(Q_n^\mathrm{b}, \Lambda_\mathrm{b}) \to 0$.

Next, we bound the first summand $\zeta_2\big(X_n, Q_n^\mathrm{b}\big)$ in (4.15) by conditioning on $I^{(n)}$. Note that, conditionally on $I^{(n)}$, the toll term $b_\mathrm{b}(n)$ is deterministic and, for the integration, denoted by $\beta = \beta\big(I^{(n)}\big)$. Defining $\mathbf{i} := (i_1, \dots, i_K)$ and denoting the distribution of $I^{(n)}$ by Υ_n, this yields

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

where, for the second inequality, we use that ζ_2 is $(2,+)$ ideal, as well as (3.7). Altogether, the estimate started in (4.15) yields

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

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

$$\Delta(n) \leq K\mathbb{E}\left[\left(\frac{I_1^{(n)}}{n}\right)^{2\lambda}\Delta\Big(I_1^{(n)}\Big)\right] + \mathrm{o}(1)\,. \tag{4.17}$$

Now, a standard argument implies $\Delta(n) \to 0$ as follows: First from (4.17) 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

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

Since $\vartheta < 1$, this implies that the sequence $(\Delta(n))_{n\geq 0}$ is bounded. We denote $\sup_{n\geq 0}\Delta(n)$ by η and $\limsup_{n\to\infty}\Delta(n)$ by ξ. For any $\varepsilon > 0$, there exists an $n_0 \geq 0$ such that $\Delta(n) \leq \xi + \varepsilon$ for all $n \geq n_0$. Hence, from (4.17) we obtain

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

With $n \to \infty$ this implies

$$\xi \leq \vartheta\,(\xi + \varepsilon)\,.$$

Since $\vartheta < 1$ and $\varepsilon > 0$ is arbitrary, this implies $\xi = 0$. Hence, as $n \to \infty$, we have $\zeta_2^{\vee}\big((X_n, Y_n), (\Lambda_{\mathrm{b}}, \Lambda_{\mathrm{w}})\big) \to 0$. Since convergence in ζ_2 implies weak convergence, this implies (4.12) as well. □

The normal limit case. Now, we discuss the normal limit case $a - c \leq (a+b)/2$, where we first consider the case with the strict inequality $a - c < (a+b)/2$. (The remaining case $a - c = (a+b)/2$ is similar with more involved expansions for the first two moments.) The formulae (4.1), (4.2) now imply

$$\mu_{\mathrm{b}}(n) = c_{\mathrm{b}} n + \mathrm{o}\big(\sqrt{n}\,\big)\,, \quad \mu_{\mathrm{w}}(n) = c_{\mathrm{w}} n + \mathrm{o}\big(\sqrt{n}\,\big) \tag{4.18}$$

with c_{b} and c_{w} as in (4.4). As usual when using 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 $\sigma_{\mathrm{w}}^2(n)$. Additionally to $bc = 0$ we exclude the case $a = c$. (In this case there is a trivial, non-random evolution of the urn). From [3], we have as $n \to \infty$:

$$\sigma_{\mathrm{b}}^2(n) = f_{\mathrm{b}} n + \mathrm{o}(n)\,, \quad \sigma_{\mathrm{w}}^2(n) = f_{\mathrm{w}} n + \mathrm{o}(n)\,, \tag{4.19}$$

with

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

We use the normalizations $X_0 := Y_0 := X_1 := Y_1 := 0$ and

$$X_n := \frac{B_n^{\mathrm{b}} - \mu_{\mathrm{b}}(n)}{\sigma_{\mathrm{b}}(n)}, \quad Y_n := \frac{B_n^{\mathrm{w}} - \mu_{\mathrm{w}}(n)}{\sigma_{\mathrm{w}}(n)}, \quad n \geq 2. \tag{4.20}$$

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

$$X_n \stackrel{d}{=} \sum_{r=1}^{a+1} \frac{\sigma_{\mathrm{b}}\big(I_r^{(n)}\big)}{\sigma_{\mathrm{b}}(n)} X^{(r)}_{I_r^{(n)}} + \sum_{r=a+2}^{K} \frac{\sigma_{\mathrm{w}}\big(I_r^{(n)}\big)}{\sigma_{\mathrm{w}}(n)} Y^{(r)}_{I_r^{(n)}} + e_{\mathrm{b}}(n)\,, \tag{4.21}$$

$$Y_n \stackrel{d}{=} \sum_{r=1}^{c} \frac{\sigma_{\mathrm{b}}\big(I_r^{(n)}\big)}{\sigma_{\mathrm{b}}(n)} X^{(r)}_{I_r^{(n)}} + \sum_{r=c+1}^{K} \frac{\sigma_{\mathrm{w}}\big(I_r^{(n)}\big)}{\sigma_{\mathrm{w}}(n)} Y^{(r)}_{I_r^{(n)}} + e_{\mathrm{w}}(n)\,, \tag{4.22}$$

with conditions on independence and identical distributions analogously to (2.2) and (2.3). We have $\|e_{\mathrm{b}}(n)\|_\infty, \|e_{\mathrm{w}}(n)\|_\infty \to 0$ since the leading linear terms in the expansions (4.18) cancel out and the error terms of order $\mathrm{o}(\sqrt{n})$ 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

$$X \stackrel{d}{=} \sum_{r=1}^{a+1} \sqrt{D_r} X^{(r)} + \sum_{r=a+2}^{K} \sqrt{D_r} Y^{(r)}, \tag{4.23}$$

$$Y \stackrel{d}{=} \sum_{r=1}^{c} \sqrt{D_r} X^{(r)} + \sum_{r=c+1}^{K} \sqrt{D_r} Y^{(r)}, \tag{4.24}$$

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 and the $Y^{(r)}$ are distributed as Y. To the map associated to the system (4.23)–(4.24) we can apply Theorem 3.2. The conditions (3.14) and (3.15) 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 4.2. *Consider the Pólya urn with replacement matrix (2.1) satisfying* $a - c < (a+b)/2$ *and* $bc > 0$. *Denote by* X_n *and* Y_n *the normalized numbers of black balls as in (4.20). Then, as* $n \to \infty$,

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

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

$$X_n \xrightarrow{d} \mathcal{N}(0,1)\,, \quad Y_n \xrightarrow{d} \mathcal{N}(0,1)\,. \tag{4.25}$$

Proof. The proof of this theorem can be done along the lines of the proof of Theorem 4.1. However, more care has to be taken in the definition of the quantities corresponding to Q_n^{b} and Q_n^{w} in (4.13) in order to assure finiteness of the ζ_3 distances. A possible choice is, for $n \geq 2$,

$$\widetilde{Q}_n^{\mathrm{b}} := \sum_{r=1}^{a+1} \mathbb{1}_{\{I_r^{(n)} \geq 2\}} \frac{\sigma_{\mathrm{b}}\big(I_r^{(n)}\big)}{\sigma_{\mathrm{b}}(n)} N_r + \sum_{r=a+2}^{K} \mathbb{1}_{\{I_r^{(n)} \geq 2\}} \frac{\sigma_{\mathrm{w}}\big(I_r^{(n)}\big)}{\sigma_{\mathrm{w}}(n)} N_r + e_{\mathrm{b}}(n)\,, \tag{4.26}$$

$$\widetilde{Q}_n^{\mathrm{w}} := \sum_{r=1}^{c} \mathbb{1}_{\{I_r^{(n)} \geq 2\}} \frac{\sigma_{\mathrm{b}}\big(I_r^{(n)}\big)}{\sigma_{\mathrm{b}}(n)} N_r + \sum_{r=c+1}^{K} \mathbb{1}_{\{I_r^{(n)} \geq 2\}} \frac{\sigma_{\mathrm{w}}\big(I_r^{(n)}\big)}{\sigma_{\mathrm{w}}(n)} N_r + e_{\mathrm{w}}(n)\,, \tag{4.27}$$

with $e_{\mathrm{b}}(n)$ and $e_{\mathrm{w}}(n)$ as in (4.21)–(4.22) and $N_1, \ldots, N_K$, $I^{(n)}$, independent, where the N_r are standard normally distributed for $r = 1, \ldots, K$. A comparison of the definition of $\widetilde{Q}_n^{\mathrm{b}}$ and $\widetilde{Q}_n^{\mathrm{w}}$ with the right hand sides of (4.21) and (4.22) and the scaling (4.20) yields that we have $\mathbb{E}\big[\widetilde{Q}_n^{\mathrm{b}}\big] = \mathbb{E}\big[\widetilde{Q}_n^{\mathrm{w}}\big] = 0$ and $\mathrm{Var}\big(\widetilde{Q}_n^{\mathrm{b}}\big) = \mathrm{Var}\big(\widetilde{Q}_n^{\mathrm{w}}\big) = 1$ for all $n \geq 2$. Obviously, we also have $\big\|\widetilde{Q}_n^{\mathrm{b}}\big\|_3, \big\|\widetilde{Q}_n^{\mathrm{w}}\big\|_3 < \infty$. Hence, ζ_3 distances between X_n, Y_n, $\widetilde{Q}_n^{\mathrm{b}}$, $\widetilde{Q}_n^{\mathrm{w}}$, and $\mathcal{N}(0,1)$ are finite for all $n \geq 2$. Denoting

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

we again start with

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

Analogously to the proof of Theorem 4.1, we obtain $\zeta_3\big(\widetilde{Q}_n^{\mathrm{b}}, \mathcal{N}(0,1)\big) \to 0$ as $n \to \infty$.

The bound for $\zeta_3\big(X_n, \widetilde{Q}_n^{\mathrm{b}}\big)$ is also analogous to the proof of Theorem 4.1, now using that ζ_3 is $(3,+)$ ideal instead of $(2,+)$ ideal. This yields

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

Then we argue as in the previous proof to obtain, analogous to (4.17),

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

From this estimate, we can deduce $\widetilde{\Delta}(n) \to 0$ as for $\Delta(n)$ in the proof of Theorem 4.1. For this, we need to use that from the expansions (4.19) and Lemma 2.1 we obtain, as $n \to \infty$, that

$$\sum_{r=1}^{a+1} \mathbb{E}\left[\left(\frac{\sigma_{\mathrm{b}}\big(I_r^{(n)}\big)}{\sigma_{\mathrm{b}}(n)}\right)^3\right] + \sum_{r=a+2}^{K} \mathbb{E}\left[\left(\frac{\sigma_{\mathrm{w}}\big(I_r^{(n)}\big)}{\sigma_{\mathrm{w}}(n)}\right)^3\right] \to \sum_{r=1}^{K} \mathbb{E}\Big[D_r^{3/2}\Big] < 1. \qquad (4.28)$$

□

Remarks. **(1)** Note that the proof of Theorem 4.2 cannot be done in the $\zeta_2^\vee$ metric since the term corresponding to (4.28) then is

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

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

(2) The case $a - c = (a + b)/2$ differs in the error terms in (4.18) which then become $\mathrm{O}(\sqrt{n})$. Since the variances in (4.19) get additional logarithmic factors, we still obtain the system (4.23)–(4.24) and the same proof technique can be applied.

(3) The condition $bc > 0$ cannot be dropped. In our approach, this would lead to degenerate systems of limit equations that do not identify limit laws. It is known that under $bc = 0$ different asymptotic behavior appears, see [16].

4.2 An urn with random replacements

As an example for an urn model with random entries in the replacement matrix R, we consider a simple model with two colors, black and white. In each step, when drawing a black ball, 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 for 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.

This type of urn has been used to model the assignment of patients to different treatment groups in clinical trials in cases when (due for example to ethical reasons) it is desirable to adapt the number of patients in each group to the efficacy of the corresponding treatment. When a new patient arrives, a ball is drawn at random from the urn and its color determines which treatment is used. The coin flips then model the success or failure of the corresponding treatment. In the course of the trial, compared to a purely random assignment, more patients are assigned to the more successfull treatment. This urn model has been studied together with generalizations in [35, 36, 34, 33, 24, 4, 5, 15].

We denote the replacement matrix by

$$R = \begin{bmatrix} F_\alpha & 1-F_\alpha \\ 1-F_\beta & F_\beta \end{bmatrix}, \tag{4.29}$$

where F_α and F_β denote Bernoulli random variables being 1 with probabilities α and β respectively, otherwise 0. The row sums of R in (4.29) are both deterministically 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}, when starting with a white ball by B_n^{w}. According to our approach in chapter 2 we obtain the recursive equation

$$B_n^{\mathrm{b}} \stackrel{d}{=} B_{I_n}^{\mathrm{b},(1)} + F_\alpha B_{J_n}^{\mathrm{b},(2)} + (1-F_\alpha)\, B_{J_n}^{\mathrm{w}}, \quad n \geq 1, \tag{4.30}$$

where $\big(B_k^{\mathrm{b},(1)}\big)_{0\leq k<n}$, $\big(B_k^{\mathrm{b},(2)}\big)_{0\leq k<n}$, $\big(B_k^{\mathrm{w}}\big)_{0\leq k<n}$, 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 $k = 0, \dots, n-1$. Furthermore, 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 when starting with one ball of each color.) Similarly, we obtain for B_n^{w} that

$$B_n^{\mathrm{w}} \stackrel{d}{=} B_{I_n}^{\mathrm{w},(1)} + F_\beta B_{J_n}^{\mathrm{w},(2)} + (1-F_\beta)\, B_{J_n}^{\mathrm{b}}, \quad n \geq 1, \tag{4.31}$$

with conditions on independence and identical distributions similar to (4.30). Together with the initial value $\big(B_0^{\mathrm{b}}, B_0^{\mathrm{w}}\big) = (1, 0)$, the system of equations (4.30)–(4.31) again defines the sequence of pairs of distributions $\left(\mathcal{L}\left(B_n^{\mathrm{b}}\right), \mathcal{L}(B_n^{\mathrm{w}})\right)_{n\geq 0}$. As a special case of Lemma 2.1, we have, for $n \to \infty$,

$$(I_n, J_n) \to (U, 1-U)\,, \tag{4.32}$$

almost surely, where U is uniformly distributed on $[0,1]$. Furthermore, for $n \geq 0$, we use the notation

$$\mu_{\mathrm{b}}(n) := \mathbb{E}\big[B_n^{\mathrm{b}}\big], \quad \mu_{\mathrm{w}}(n) := \mathbb{E}\big[B_n^{\mathrm{w}}\big]. \tag{4.33}$$

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

Lemma 4.3. *For $\mu_{\mathrm{b}}(n)$ and $\mu_{\mathrm{w}}(n)$ as in (4.33) with $0 < \alpha, \beta < 1$, we have*

$$\mu_{\mathrm{b}}(n) = \frac{1-\beta}{2-\alpha-\beta}\,n + \frac{1-\alpha}{2-\alpha-\beta}\,\frac{\Gamma(n+\alpha+\beta)}{\Gamma(\alpha+\beta)\,\Gamma(n+1)} + \frac{1-\beta}{2-\alpha-\beta}, \tag{4.34}$$

$$\mu_{\mathrm{w}}(n) = \frac{1-\beta}{2-\alpha-\beta}\,n - \frac{1-\beta}{2-\alpha-\beta}\,\frac{\Gamma(n+\alpha+\beta)}{\Gamma(\alpha+\beta)\,\Gamma(n+1)} + \frac{1-\beta}{2-\alpha-\beta}. \tag{4.35}$$

Proof. An elementary proof is based on matrix diagonalization and can be done along the lines of the proof of Lemma 4.7 below. □

As in the example in section 4.1, we have two different types of limit laws, with normal limit for $\alpha+\beta \leq 3/2$ and with non-normal limit for $\alpha+\beta > 3/2$.

The non-normal limit case. We assume that $\lambda := \alpha+\beta-1 > 1/2$. From Lemma 4.3, we obtain asymptotic expressions for the expectation, as $n \to \infty$,

$$\mu_{\mathrm{b}}(n) = c'_{\mathrm{b}} n + d'_{\mathrm{b}} n^{\lambda} + \mathrm{o}\Big(n^{\lambda}\Big)\,,$$

$$\mu_{\mathrm{w}}(n) = c'_{\mathrm{w}} n + d'_{\mathrm{w}} n^{\lambda} + \mathrm{o}\Big(n^{\lambda}\Big)\,,$$

with constants

$$c'_{\mathrm{b}} = c'_{\mathrm{w}} = \frac{1-\beta}{1-\lambda}, \quad d'_{\mathrm{b}} = \frac{1-\alpha}{(1-\lambda)\,\Gamma(\lambda+1)}, \quad d'_{\mathrm{w}} = -\frac{1-\beta}{(1-\lambda)\,\Gamma(\lambda+1)}. \tag{4.36}$$

We use the normalizations $X_0 := Y_0 := 0$ and

$$X_n := \frac{B_n^{\mathrm{b}} - \mu_{\mathrm{b}}(n)}{n^{\lambda}}, \quad Y_n := \frac{B_n^{\mathrm{w}} - \mu_{\mathrm{w}}(n)}{n^{\lambda}}, \quad n \geq 1. \tag{4.37}$$

As in the non-normal case of the example in section 4.1, it is sufficient to use the order of the error term of the mean for the scaling. From (4.30)–(4.31) we obtain, for $n \geq 1$,

$$X_n \stackrel{d}{=} \left(\frac{I_n}{n}\right)^{\lambda} X^{(1)}_{I_n} + F_\alpha \left(\frac{J_n}{n}\right)^{\lambda} X^{(2)}_{J_n} + (1 - F_\alpha) \left(\frac{J_n}{n}\right)^{\lambda} Y_{J_n} + b'_{\mathrm{b}}(n)\,, \tag{4.38}$$

$$Y_n \stackrel{d}{=} \left(\frac{I_n}{n}\right)^{\lambda} Y^{(1)}_{I_n} + F_\beta \left(\frac{J_n}{n}\right)^{\lambda} Y^{(2)}_{J_n} + (1 - F_\beta) \left(\frac{J_n}{n}\right)^{\lambda} X_{J_n} + b'_{\mathrm{w}}(n)\,, \tag{4.39}$$

where

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

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

with conditions on independence and identical distributions analogous to (4.30)–(4.31). In view of (4.32), this suggests for limits X and Y of X_n and Y_n that

$$X \stackrel{d}{=} U^{\lambda} X^{(1)} + F_\alpha \left(1 - U\right)^{\lambda} X^{(2)} + \left(1 - F_\alpha\right) \left(1 - U\right)^{\lambda} Y^{(1)} + b'_{\mathrm{b}}, \tag{4.40}$$

$$Y \stackrel{d}{=} U^{\lambda} Y^{(1)} + F_\beta \left(1 - U\right)^{\lambda} Y^{(2)} + \left(1 - F_\beta\right) \left(1 - U\right)^{\lambda} X^{(1)} + b'_{\mathrm{w}}, \tag{4.41}$$

with

$$b'_{\mathrm{b}} = d'_{\mathrm{b}} \left(U^{\lambda} + F_\alpha (1 - U)^{\lambda} - 1\right) + d'_{\mathrm{w}} \left(1 - F_\alpha\right) \left(1 - U\right)^{\lambda},$$

$$b'_{\mathrm{w}} = d'_{\mathrm{w}} \left(U^{\lambda} + F_\beta (1 - U)^{\lambda} - 1\right) + d'_{\mathrm{b}} \left(1 - F_\beta\right) \left(1 - U\right)^{\lambda},$$

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

To check that Theorem 3.1 can be applied to the map associated to the system (4.40)–(4.41), first note that the form of d'_{b} and d'_{w} in (4.36) implies that $\mathbb{E}\left[b'_{\mathrm{b}}\right] = \mathbb{E}\left[b'_{\mathrm{w}}\right] = 0$. To check condition (3.13), note that we have

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

since $\lambda > 1/2$. Analogously, we have

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

Together, this verifies condition (3.13). Hence, Theorem 3.1 can be applied and yields that the system (4.40)–(4.41) has a unique fixed-point $\big(\mathcal{L}(\Lambda'_{\mathrm{b}})\,,\mathcal{L}(\Lambda'_{\mathrm{w}})\big)$ in $\mathcal{M}_2^{\mathbb{R}}(0) \times \mathcal{M}_2^{\mathbb{R}}(0)$.

Theorem 4.4. *Consider the Pólya urn with random replacement matrix (4.29) with $\alpha, \beta \in (0,1)$ and $\alpha+\beta > 3/2$. Furthermore, let X_n and Y_n be the normalized numbers of black balls after n steps as in (4.37). We denote by $\big(\mathcal{L}(\Lambda'_{\mathrm{b}})\,,\mathcal{L}(\Lambda'_{\mathrm{w}})\big)$ the unique solution of* (4.40)–(4.41) *in $\mathcal{M}_2^{\mathbb{R}}(0) \times \mathcal{M}_2^{\mathbb{R}}(0)$. Then, as $n \to \infty$,*

$$X_n \xrightarrow{d} \Lambda'_{\mathrm{b}}, \quad Y_n \xrightarrow{d} \Lambda'_{\mathrm{w}}.$$

Proof. Analogous to the proof of Theorem 4.1. □

The normal limit case. Now, we discuss the normal limit case $\lambda := \alpha+\beta-1 \le 1/2$. We first assume $\lambda := \alpha+\beta-1 < 1/2$. The expansions from Lemma 4.3 now imply, as $n \to \infty$

$$\mu_{\mathrm{b}}(n) = c_{\mathrm{b}} n + \mathrm{o}\big(\sqrt{n}\,\big)\,, \quad \mu_{\mathrm{w}}(n) = c_{\mathrm{w}} n + \mathrm{o}\big(\sqrt{n}\,\big) \tag{4.42}$$

with c_{b} and c_{w} given in (4.36). As in the normal limit cases in the examples in section 4.1, we first need asymptotic expressions for the variances. We denote the variances of B_n^{b} and B_n^{w} by $\hat\sigma_{\mathrm{b}}^2(n)$ and $\hat\sigma_{\mathrm{w}}^2(n)$ respectively. These can be obtained from a result of Matthews and Rosenberger [24] for the number of draws of each color as follows:

Lemma 4.5. *For the variances of B_n^{b} and B_n^{w}, we have, as $n \to \infty$,*

$$\hat\sigma_{\mathrm{b}}^2(n) = f'_{\mathrm{b}} n + \mathrm{o}(n)\,, \quad \hat\sigma_{\mathrm{w}}^2(n) = f'_{\mathrm{w}} n + \mathrm{o}(n)\,, \tag{4.43}$$

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. For the present urn model, Matthews and Rosenberger [24] study the number N_n of draws, within the first n draws, in which a black ball is drawn. Starting with one black ball, it is established there that, as $n \to \infty$,

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

For each black ball in the urn, exactly one of the following three statements is true: It may be either the first ball, or has been added after drawing a black ball and having success in tossing the corresponding coin, or added after drawing a white ball and having no success in tossing the coin. Therefore, we can directly link N_n to B_n^{b}: Denoting the coin flips after drawing black balls by $\big(F_j^{\mathrm{b}}\big)_{1\le j\le N_n}$, the coin flips after drawing white balls by $\big(F_j^{\mathrm{w}}\big)_{1\le j\le (n-N_n)}$, we have

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

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

$$\hat\sigma_b^2(n) = \mathbb{E}\Big[\mathrm{Var}\big(B_n^{\mathrm{b}} \mid N_n\big)\Big] + \mathrm{Var}\Big(\mathbb{E}\big[B_n^{\mathrm{b}} \mid N_n\big]\Big)$$
$$= \frac{(1-\alpha)\,(1-\beta)}{(1-\lambda)^2}\Big(\frac{1}{1-2\lambda} - 2\lambda\,(1+\lambda)\Big)n + \mathrm{o}(n)\,.$$

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

We use the normalizations $X_0 := Y_0 := 0$ and

$$X_n := \frac{B_n^{\mathrm{b}} - \mu_{\mathrm{b}}(n)}{\hat\sigma_{\mathrm{b}}(n)}, \quad Y_n := \frac{B_n^{\mathrm{w}} - \mu_{\mathrm{w}}(n)}{\hat\sigma_{\mathrm{w}}(n)}, \quad n \ge 1. \tag{4.44}$$

From the system (4.30)–(4.31), we obtain for the scaled quantities X_n, Y_n the system, for $n \ge 1$,

$$X_n \stackrel{d}{=} \frac{\hat\sigma_{\mathrm{b}}(I_n)}{\hat\sigma_{\mathrm{b}}(n)} X_{I_n}^{(1)} + F_\alpha \frac{\hat\sigma_{\mathrm{b}}(J_n)}{\hat\sigma_{\mathrm{b}}(n)} X_{J_n}^{(2)} + (1-F_\alpha)\,\frac{\hat\sigma_{\mathrm{w}}(J_n)}{\hat\sigma_{\mathrm{w}}(n)} Y_{J_n} + e_{\mathrm{b}}'(n)\,,$$

$$Y_n \stackrel{d}{=} \frac{\hat\sigma_{\mathrm{w}}(I_n)}{\hat\sigma_{\mathrm{w}}(n)} Y_{I_n}^{(1)} + F_\beta \frac{\hat\sigma_{\mathrm{w}}(J_n)}{\hat\sigma_{\mathrm{w}}(n)} Y_{J_n}^{(2)} + (1-F_\beta)\,\frac{\hat\sigma_{\mathrm{b}}(J_n)}{\hat\sigma_{\mathrm{b}}(n)} X_{J_n} + e_{\mathrm{w}}'(n)\,,$$

with conditions on independence and identical distributions analogous to (4.30)–(4.31). We have $\left\|e'_{\mathrm{b}}(n)\right\|_\infty, \left\|e'_{\mathrm{w}}(n)\right\|_\infty \to 0$, since the leading linear terms in the expansions (4.42) cancel out and the error terms of order $\mathrm{o}(\sqrt{n}\,)$ are asymptotically eliminated by the scaling of order $1/\sqrt{n}$. In view of (4.32), this suggests for limits X and Y of X_n and Y_n respectively

$$X \stackrel{d}{=} \sqrt{U}X^{(1)} + F_\alpha\sqrt{1-U}X^{(2)} + (1-F_\alpha)\sqrt{1-U}Y^{(1)}, \tag{4.45}$$

$$Y \stackrel{d}{=} \sqrt{U}Y^{(1)} + F_\beta\sqrt{1-U}Y^{(2)} + (1-F_\beta)\sqrt{1-U}X^{(1)}, \tag{4.46}$$

where $X^{(1)}$, $X^{(2)}$, $Y^{(1)}$, $Y^{(2)}$ and U are independent and $X^{(1)}$, $X^{(2)}$ are distributed as X whereas $Y^{(1)}$, $Y^{(2)}$ are distributed as Y. We can apply Theorem 3.2 to the map associated to the system (4.45)–(4.46), because the conditions (3.14) and (3.15) are trivially satisfied. Hence, $\big(\mathcal{N}(0,1)\,,\mathcal{N}(0,1)\big)$ 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 4.6. *Consider the Pólya urn with random replacement matrix (4.29) with* $\alpha, \beta \in (0,1)$ *and* $\alpha + \beta < 3/2$. *Furthermore, denote by* X_n *and* Y_n *the normalized numbers of black balls as in (4.44). Then, as* $n \to \infty$,

$$X_n \xrightarrow{d} \mathcal{N}(0,1)\,, \quad Y_n \xrightarrow{d} \mathcal{N}(0,1)\,. \tag{4.47}$$

Proof. Analogous to the proof of Theorem 4.2. □

Remark. The case $\alpha+\beta = 3/2$ differs in the error terms in (4.42) which then become $\mathrm{O}(\sqrt{n})$. Since the variances in (4.43) get additional logarithmic factors (cf. [24]), we still obtain the system (4.45)–(4.46) and our proof technique still applies.

4.3 Cyclic urns

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

$$R = \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \vdots & & \ddots & \ddots & 0 \\ 0 & & & \ddots & 1 \\ 1 & 0 & \cdots & \cdots & 0 \end{bmatrix}. \tag{4.48}$$

We denote by $R_n^{[j]}$ 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

$$\begin{aligned} R_n^{[1]} &\stackrel{d}{=} R_{I_n}^{[1]} + R_{J_n}^{[2]}, \\ R_n^{[2]} &\stackrel{d}{=} R_{I_n}^{[2]} + R_{J_n}^{[3]}, \\ &\vdots \\ R_n^{[m]} &\stackrel{d}{=} R_{I_n}^{[m]} + R_{J_n}^{[1]}, \end{aligned} \tag{4.49}$$

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

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

$$\omega := \exp\left(\frac{2\pi \mathrm{i}}{m}\right) =: \lambda + \mathrm{i}\mu \tag{4.50}$$

with $\lambda, \mu \in \mathbb{R}$. Note that for $2 \leq m \leq 6$, we have $\lambda \leq 1/2$, while for $m \geq 7$, we have $\lambda > 1/2$. Asymptotic expressions for the means of $R_n^{[j]}$ can be found (together with further analysis) in [14, 15, 29]. To keep this section self-contained, we give an exact formula for later use:

Lemma 4.7. *Let $R_n^{[j]}$ be the number of balls of color 1 after n draws in a cyclic urn with $m \geq 2$ colors, starting with one ball of color j. Then, with $\omega = \omega_m$ as in (4.50), we have*

$$\mathbb{E}\Big[R_n^{[j]}\Big] = \frac{n+1}{m} + \frac{1}{m} \sum_{\substack{k\in\{1,\dots,m-1\} \\ k \neq m/2}} \frac{\Gamma(n+1+\omega^k)}{\Gamma(n+1)\,\Gamma(\omega^k+1)} \omega^{k(j-1)}. \tag{4.51}$$

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

$$\mathbb{E}\left[R_n^{[j]}\right] = \frac{1}{m}\, n + \Re\big(\kappa_j n^{\mathrm{i}\mu}\big)\, n^{\lambda} + \mathrm{o}\Big(n^{\lambda}\Big), \qquad \kappa_j := \frac{2\omega^{j-1}}{m\Gamma(\omega+1)}. \tag{4.52}$$

Proof. Using the system (4.49), we obtain by conditioning on I_n, for any $1 \leq j \leq m$,

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

Using the column vector $R_n := \big(R_n^{[1]}, \dots, R_n^{[m]}\big)^t$, the replacement matrix R in (4.48), and the identity matrix Id_m, we can rewrite this to get

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

The eigenvalues of the replacement matrix R are all m-th roots of unity $\omega_\ell := \omega^\ell$, $\ell = 1, \dots, m$. and a possible eigenbasis is $v_\ell := \frac{1}{m}\left(\omega_\ell^0, \dots, \omega_\ell^{m-1}\right)^t$, $\ell = 1, \dots, m$. Decomposing the mapping induced by R into the projections π_{v_ℓ} onto the respective eigenspaces, we obtain

$$\begin{aligned}
\prod_{k=1}^{n} \left(\mathrm{Id}_m + \frac{1}{k}R\right) &= \sum_{l=1}^{m} \prod_{k=1}^{n} \left(1 + \frac{1}{k}\omega_l\right) \pi_{v_l} \\
&= (n+1)\; \pi_{v_m} + \sum_{\substack{l\in\{1,\dots,m-1\} \\ l \neq m/2}} \frac{\Gamma(n+1+\omega_l)}{\Gamma(\omega_l+1)\Gamma(n+1)} \pi_{v_l}.
\end{aligned}$$

Moreover, $\pi_{v_j}\big(\mathbb{E}[R_0]\big) = v_j$ and $v_m = \frac{1}{m}(1,\dots,1)$, hence the j-th component of the latter display implies (4.51).

The asymptotic expansion in (4.52) can now directly be read off: Note that the roots of unity come in conjugate pairs $\overline{\omega_\ell} = \omega_{m-\ell}$. If m is even, $\omega_{m/2} = \overline{\omega_{m/2}} = -1$, otherwise only $\omega_m = 1$ is real. Combining the summands for such conjugate pairs and using that $\Gamma(\overline{z}) = \overline{\Gamma(z)}$, we obtain the terms

$$\frac{\Gamma(n+1+\omega_\ell)\,\omega_\ell^{j-1}}{\Gamma(n+1)\,\Gamma(\omega_\ell+1)} + \frac{\Gamma(n+1+\overline{\omega_\ell})\,\overline{\omega_\ell}^{j-1}}{\Gamma(n+1)\,\Gamma(\overline{\omega_\ell}+1)} = 2\,\Re\left(\frac{\omega_\ell^{j-1}\,\Gamma(n+1+\omega_\ell)}{\Gamma(\omega_\ell+1)\,\Gamma(n+1)}\right). \tag{4.53}$$

Therefore, the expectation can be rewritten as

$$\mathbb{E}\Big[R_n^{[j]}\Big] = \frac{n+1}{m} + \frac{2}{m}\sum_{l=1}^{\lfloor (m-1)/2\rfloor} \Re\left(\frac{\Gamma(n+1+\omega_\ell)}{\Gamma(n+1)\,\Gamma(\omega_\ell+1)}\,\omega_\ell^{j-1}\right) + \mathbb{1}_{\{m \text{ even}\}}\frac{(-1)^{j-1}}{mn}.$$

By Stirling's approximation, the asymptotic growth order of the term in (4.53) is $\Re(n^{\omega_\ell})$, hence the dominant asymptotic term is the one for the conjugate pair with largest real part, ω and ω_{m-1}. This implies (4.52) for $m > 4$. For $m = 3, 4$ the periodic term is $o(1)$ respectively $O(1)$, whereas for $m = 2$ there is no periodic fluctuation. □

We do not discuss limit laws for the cases $2 \le m \le 6$ in detail. They lead to asymptotic normality as has been shown with different proofs in Janson [14] and Janson [15, Example 7.9]. These cases can be covered by our approach similarly to the normal cases in sections 4.1 and 4.2. For $2 \le m \le 6$, 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 3.2 applies.

We now assume $m \geq 7$. In particular, we have the asymptotic expansion (4.52) of the means of the $R_n^{[j]}$ with $\lambda > 1/2$. We define the normalizations $X_0^{[j]} = 0$ and

$$X_n^{[j]} := \frac{R_n^{[j]} - \frac{1}{m} n}{n^\lambda}, \qquad n \geq 1. \tag{4.54}$$

Hence, we obtain for $n \geq 1$ the system

$$X_n^{[1]} \stackrel{d}{=} \left(\frac{I_n}{n}\right)^\lambda X_{I_n}^{[1]} + \left(\frac{J_n}{n}\right)^\lambda X_{J_n}^{[2]} - \frac{1}{m\, n^\lambda},$$

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

$$\vdots$$

$$X_n^{[m]} \stackrel{d}{=} \left(\frac{I_n}{n}\right)^\lambda X_{I_n}^{[m]} + \left(\frac{J_n}{n}\right)^\lambda X_{J_n}^{[1]} - \frac{1}{m\, n^\lambda},$$

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

$$X^{[1]} \stackrel{d}{=} U^\omega X^{[1]} + (1-U)^\omega X^{[2]},$$

$$X^{[2]} \stackrel{d}{=} U^\omega X^{[2]} + (1-U)^\omega X^{[3]},$$

$$\vdots$$

$$X^{[m]} \stackrel{d}{=} U^\omega X^{[m]} + (1-U)^\omega X^{[1]}.$$

Since ω is complex, this now has to be considered as a system to solve for distributions $\mathcal{L}\big(X^{[1]}\big), \ldots, \mathcal{L}\big(X^{[m]}\big)$ on the complex plane $\mathbb{C}$. The corresponding map $\bar{T}$ is a special case of T' in (3.12):

$$\bar{T} : \left(\mathcal{M}^{\mathbb{C}}\right)^{\times m} \to \left(\mathcal{M}^{\mathbb{C}}\right)^{\times m}$$

$$(\mu_1, \ldots, \mu_m) \mapsto \Big(\bar{T}_1(\mu_1, \ldots, \mu_m), \ldots, \bar{T}_m(\mu_1, \ldots, \mu_m)\Big)$$

$$\bar{T}_j(\mu_1, \ldots, \mu_m) := \mathcal{L}\Big(U^\omega V^{[j]} + (1-U)^\omega V^{[j+1]}\Big) \tag{4.55}$$

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

Lemma 4.8. *Let $m \geq 7$. 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 3.3: First note that condition (3.16) for $\bar{T}$ in (4.55) is

$$\mathbb{E}\big[U^\omega\big]\kappa_j + \mathbb{E}\big[(1-U)^\omega\big]\kappa_{j+1} = \kappa_j, \quad j = 1, \ldots, m, \tag{4.56}$$

with $\kappa_{m+1} \coloneqq \kappa_1$. Since $\mathbb{E}[U^\omega] = \mathbb{E}\big[(1-U)^\omega\big] = (1+\omega)^{-1}$ and $\kappa_{j+1} = \omega\kappa_j$, we find that (4.56) is satisfied. Condition (3.17) for $\bar{T}$ is

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

The left-hand side is equal to $2/(1+2\lambda)$ and, since $m \geq 7$, we have $\lambda > 1/2$. Hence, Theorem 3.3 applies and implies the assertion. □

The fixed-point in Lemma 4.8 has a particularly simple structure as follows:

Lemma 4.9. *Let $m \geq 7$ and $\big(\mathcal{L}(\Lambda^{[1]}), \ldots, \mathcal{L}(\Lambda^{[m]})\big)$ be the unique fixed-point in Lemma 4.8. Furthermore, let $\mathcal{L}(\Lambda)$ be the (unique)fixed-point of*

$$X \overset{d}{=} U^\omega X + \omega\,(1-U)^\omega\, X' \quad \text{in} \quad \mathcal{M}_2^{\mathbb{C}}\left(\frac{2}{m\,\Gamma(\omega+1)}\right), \tag{4.57}$$

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

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

Proof. We abbreviate $\gamma \coloneqq 2/\big(m\,\Gamma(\omega+1)\big)$. 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 (4.57) maps $\mathcal{M}_2^{\mathbb{C}}(\gamma)$ into itself. The argument of the proof of Theorem 3.3 implies that this map is a contraction on $\big(\mathcal{M}_2^{\mathbb{C}}(\gamma)\,,\ell_2\big)$. Hence it has a unique fixed point $\mathcal{L}(\Lambda)$. We have

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

and, by plugging into (4.55), we find that this vector is a fixed-point of $\bar{T}$. Since, by Lemma 4.8, 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 behavior in the following theorem has already been shown almost surely by martingale methods in [29, Section 4.2], see also [15, Theorem 3.24]. Our contraction approach adds the characterization of $\mathcal{L}(\Lambda)$ as the fixed-point in (4.57). The proof is based on the complex version of the ℓ_2 metric and resembles ideas from Fill and Kapur [12], see also [17, Theorem 5.3].

Theorem 4.10. *Let $m \geq 7$ and $X_n^{[j]}$ as in (4.54) and $\mathcal{L}(\Lambda)$ be the unique fixed-point in Lemma 4.9. Then, for all $j = 1, \dots, m$, we have*

$$\ell_2\Big(X_n^{[j]}, \Re\Big(e^{\mathrm{i}\left(\mu \log(n) + 2\pi \frac{j-1}{m}\right)}\Lambda\Big)\Big) \to 0 \quad (n \to \infty). \tag{4.58}$$

Proof. Let $\Lambda^{[1]}, \dots, \Lambda^{[m]}$ be independent random variables such that the vector of their distributions $\left(\mathcal{L}\left(\Lambda^{[1]}\right), \dots, \mathcal{L}\left(\Lambda^{[m]}\right)\right)$ is the unique fixed-point as in Lemma 4.8. Set $\Lambda^{[m+1]} := \Lambda^{[1]}$ and note that for the random variable within the real part in (4.58) with Lemma 4.9, we have

$$e^{\mathrm{i}\left(\mu \log(n) + 2\pi \frac{j-1}{m}\right)}\Lambda = n^{\mathrm{i}\mu}\omega^{j-1}\Lambda \stackrel{d}{=} n^{\mathrm{i}\mu}\Lambda^{[j]}. \tag{4.59}$$

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

$$\Re\Big(n^{\mathrm{i}\mu}\Lambda^{[j]}\Big) \stackrel{d}{=} \Re\Big(n^{\mathrm{i}\mu}U^{\omega}\Lambda^{[j]}\Big) + \Re\Big(n^{\mathrm{i}\mu}\left(1-U\right)^{\omega}\Lambda^{[j+1]}\Big) \tag{4.60}$$

for all $j = 1, \dots, m$ and $n \geq 0$. Note that here and in the following we silently identify $m+1$ and 1.

Now, we assume that for all $n \geq 1$, all $X_n^{[j]}$ and $\Lambda^{[j]}$ for $1 \leq j \leq m$, I_n, and U appearing in (4.54) and (4.55) are defined on one probability space such that $\left(X_n^{[j]}, \Re\left(n^{\mathrm{i}\mu}\Lambda^{[j]}\right)\right)$

are optimal ℓ_2-couplings for all $n \geq 0$ and all $1 \leq j \leq m$ and such that $I_n = \lfloor nU \rfloor$. Then we have

$$\begin{aligned}
\Delta_j(n) &:= \ell_2\Big(X_n^{[j]}, \Re\big(n^{\mathrm{i}\mu}\Lambda^{[j]}\big)\Big) \\
&= \ell_2\left(\left(\frac{I_n}{n}\right)^{\lambda} X_{I_n}^{[j]} + \left(\frac{J_n}{n}\right)^{\lambda} X_{J_n}^{[j+1]} - \frac{1}{mn^{\lambda}}, \Re\Big(n^{\mathrm{i}\mu}U^{\omega}\Lambda^{[j]}\Big) + \Re\Big(n^{\mathrm{i}\mu}(1-U)^{\omega}\Lambda^{[j+1]}\Big)\right) \\
&\leq \left\| \left\{ \left(\frac{I_n}{n}\right)^{\lambda} X_{I_n}^{[j]} - \Re\left(\frac{I_n^{\omega}}{n^{\lambda}}\Lambda^{[j]}\right)\right\} + \left\{ \left(\frac{J_n}{n}\right)^{\lambda} X_{J_n}^{[j+1]} - \Re\left(\frac{J_n^{\omega}}{n^{\lambda}}\Lambda^{[j+1]}\right)\right\} \right\|_2 \\
&\quad + \left\| \Re\left(\frac{I_n^{\omega}}{n^{\lambda}}\Lambda^{[j]}\right) - \Re\Big(n^{\mathrm{i}\mu}U^{\omega}\Lambda^{[j]}\Big) \right\|_2 + \left\| \Re\left(\frac{J_n^{\omega}}{n^{\lambda}}\Lambda^{[j+1]}\right) - \Re\Big(n^{\mathrm{i}\mu}U^{\omega}\Lambda^{[j+1]}\Big) \right\|_2 \\
&\quad + \frac{1}{mn^{\lambda}} \\
&=: S_1 + S_2 + S_3 + \frac{1}{mn^{\lambda}}. \qquad (4.61)
\end{aligned}$$

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

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

Then we have

$$S_1^2 = \mathbb{E}\left[\left(W_n^{[j]}\right)^2\right] + \mathbb{E}\left[\left(W_n^{[j+1]}\right)^2\right] + 2\,\mathbb{E}\left[W_n^{[j]}W_n^{[j+1]}\right]. \qquad (4.62)$$

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

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

Analogously, we have

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

To bound the mixed term in (4.62), note that by expansion (4.52) and normalization (4.54) we have $\mathbb{E}\big[X_n^{[j]}\big] = \Re\big(\kappa_j n^{\mathrm{i}\mu}\big) + r_j(n)$, 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 $\mathbb{E}\big[\Lambda^{[j]}\big] = \kappa_j$, this implies $\mathbb{E}\big[W_n^{[j]}\big] = \mathbb{E}\big[(I_n/n)^{\lambda}\, r_j(I_n)\big]$ and

$$\mathbb{E}\left[W_n^{[j]} W_n^{[j+1]}\right] = \mathbb{E}\left[\left(\frac{I_n}{n}\frac{J_n}{n}\right)^{\lambda} r_j(I_n)\; r_{j+1}(J_n)\right]. \tag{4.63}$$

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

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

Hence, we obtain that the mixed term (4.63) tends to zero.

Altogether, we obtain from (4.61), as $n \to \infty$, that

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

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

$$\Delta(n) \coloneqq \max_{1 \leq j \leq m} \Delta_j(n)\,.$$

Hence, we have

$$\Delta(n) \leq \left\{ 2\,\mathbb{E}\left[\left(\frac{I_n}{n}\right)^{2\lambda} \Delta^2(I_n) \right] + \mathrm{o}(1) \right\}^{1/2} + \mathrm{o}(1)\,. \tag{4.64}$$

Now, we can obtain $\Delta(n) \to 0$ in the same way as in the proof of Theorem 4.1: First, from (4.64), we obtain with $I_n/n \to U$ almost surely that

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

Since $\lambda > 1/2$, this implies that the sequence $(\Delta(n))_{n \geq 0}$ is bounded. We denote $\eta \coloneqq \sup_{n \geq 0} \Delta(n)$ and $\xi \coloneqq \limsup_{n \to \infty} \Delta(n)$. For any $\varepsilon > 0$, there exists an $n_0 \geq 0$ such that $\Delta(n) \leq \xi + \varepsilon$ for all $n \geq n_0$. Hence, from (4.64) we obtain

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

With $n \to \infty$, this implies

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

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

5 Two-dimensional recursion

Our first approach to showing convergence for the urn models was to state a recurrence relation for a vector (instead of the system of equations used in the last chapters) and try to use the general convergence theorems in Neininger [25] and Neininger and Rüschendorf [26]. In this chapter, we will discuss the problems we encountered when trying to use this approach for the example of the urn with random replacements described in section 4.2. We also state an extension of the above theorem which enables us to prove convergence for this example. We achieve this by changing the norm on $\mathbb{R}^d$, replacing the Euclidean norm by a p-norm for appropriate $p \in [1, \infty]$.

Neininger and Rüschendorf [27] discusses applications of the multidimensional approach and, amongst other things, pros and cons of using ℓ_2 or Zolotarev metric ζ_2 for the non-normal limit cases. Our extension of their general limit theorem, especially when using the supremum norm, removes both major disadvantages of the ζ_2-variant: we can weaken the condition on the expectation of the operator norm of the coefficients even below what is needed for ℓ_2 convergence and at the same time, the new condition is often even easier to check, as for any nonnegative matrix, the operator norm with respect to the supremum norm is just the maximum of the row sums of the matrix.

In section 5.1, we will first describe the two-dimensional model for the example of section 4.2. We will then give an account of the result of Neininger and Rüschendorf [26, Thm 4.1] and show that its direct application in this case does not ensure convergence for all relevant values of α and β. In section 5.3, we show that using a p-Norm ($p \in [1, \infty]$) on $\mathbb{R}^d$ as underlying norm for the definition of the Zolotarev metric allows us to extend the scope of the theorem. In section 5.4, we use this result to show convergence in the non-normal case and give an outline of the proof for the other cases.

5.1 Two-dimensional recursive equation

In the following, we will work with bivariate recurrences for the random vector $B_n := \big(B_n^{\mathrm{b}}, B_n^{\mathrm{w}}\big)^t$. Note that in the previous discussion the random variable 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 $\big(B_n^{\mathrm{b}}, B_n^{\mathrm{w}}\big)$ are determined by the urn process and we have the choice of a joint distribution for $\big(B_n^{\mathrm{b}}, B_n^{\mathrm{w}}\big)$ respecting these marginals. We could keep the components independent or choose appropriate couplings. We choose a version with a coupling defined recursively by $B_0 := (1,0)^t$ and, for $n \geq 1$,

$$B_n \overset{d}{:=} \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} B_{I_n} + \begin{pmatrix} F_\alpha & 1 - F_\alpha \\ 1 - F_\beta & F_\beta \end{pmatrix} B'_{J_n} \tag{5.1}$$

where $(B_k)_{0 \leq k < n}$, $(B'_k)_{0 \leq k < n}$, (F_α, F_β), and I_n are independent and $B_n \overset{d}{=} B'_n$ for all $n \geq 0$. As in section 4.2, I_n is uniformly distributed on $\{0, \ldots, n-1\}$ 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, otherwise 0. Note that for any joint distribution of (F_α, F_β), definition (5.1) leads to a sequence $(B_n)_{n \geq 1}$ with correct marginals $\mathcal{L}\big(B_n^{\mathrm{b}}\big)$ and $\mathcal{L}(B_n^{\mathrm{w}})$. A particular joint distribution will be chosen later.

5.2 Multivariate convergence theorem and application

For random vectors satisfying multivariate recurrences as the one stated above, Neininger [25] and Neininger and Rüschendorf [26] state general transfer theorems of the form that appropriate convergence of the coefficients (after scaling) together with some technical requirements implies weak convergence of the random vector to a limit distribution which can be characterized as the unique solution of a distributional fixed-point equation. We will discuss here the problems we encountered when trying to apply the transfer theorem of Neininger and Rüschendorf [26, Thm 4.1] to the example of the urn with random replacements described in section 4.2. For this we first give an account of the general setting and assertion of the theorem, restricted to the cases relevant here. We then investigate the requirements of the theorem for the mentioned example.

5.2.1 Spaces of distributions and metrics

To work in the multidimensional setting we first have to specify the space of distributions we are working in. This is similar to section 3.1, but we now directly consider (multidimensional) probability distributions on $\mathbb{R}^d$ and corresponding subsets thereof.

Spaces. Instead of the cartesian product $\left(\mathcal{M}^{\mathbb{R}}\right)^{\times d}$, which we used in the first chapters, we now work in $\mathcal{M}^{\mathbb{R}^d}$, the space of all probability distributions on $\mathbb{R}^d$ with the Borel σ-field, and its subspaces

$$\mathcal{M}_s^{\mathbb{R}^d} := \left\{\mathcal{L}(X) \in \mathcal{M}^{\mathbb{R}^d} \mid \mathbb{E}\left[\|X\|^s\right] < \infty\right\}, \quad s > 0, \tag{5.2}$$

$$\mathcal{M}_s^{\mathbb{R}^d}(\mu) := \left\{\mathcal{L}(X) \in \mathcal{M}_s^{\mathbb{R}^d} \mid \mathbb{E}[X] = \mu\right\}, \quad s \geq 1,\ \mu \in \mathbb{R}^d, \tag{5.3}$$

$$\mathcal{M}_s^{\mathbb{R}^d}(\mu, C) := \left\{\mathcal{L}(X) \in \mathcal{M}_s^{\mathbb{R}^d}(\mu) \mid \mathrm{Cov}(X) = C\right\}, \quad s \geq 2,\ \mu \in \mathbb{R}^d, \tag{5.4}$$

where C is a symmetric positive semidefinite $d \times d$ matrix.

Zolotarev metric. On $\mathcal{M}^{\mathbb{R}^d}$, the Zolotarev metric can be defined similar to (3.4). We first define the version of the Zolotarev metric ζ_s used by Neininger and Rüschendorf [26],

$$\zeta_s(X, Y) := \zeta_s\big(\mathcal{L}(X), \mathcal{L}(Y)\big) := \sup_{f \in \mathcal{F}_s} \left|\mathbb{E}[f(X) - f(Y)]\right|, \tag{5.5}$$

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

$$\mathcal{F}_s := \left\{ f \in C^m\left(\mathbb{R}^d, \mathbb{R}\right) \;\middle|\; \forall\, x, y \in \mathbb{R}^d :\ \left\| f^{(m)}(x) - f^{(m)}(y) \right\|_{\mathrm{op}} \leq \|x - y\|^\alpha \right\}. \tag{5.6}$$

The properties of ζ_s cited in section 3.1 also hold for the multidimensional setting, including $(s, +)$-ideality of ζ_s, completeness of the corresponding spaces and the implication of weak convergence on $\mathbb{R}^d$, see Neininger and Rüschendorf [26] and Drmota et al. [10, Theorem 5.1].

5.2.2 General Transfer Theorem

Neininger and Rüschendorf [26, Thm 4.1] gives a general limit theorem for random vectors satisfying certain recurrence relations. Their theorem states requirements on the coefficients of the recurrence relation and implies weak convergence. For simplicity, we give a restriction of the theorem to the cases relevant here (two-dimensional, two summands, and no toll term).

Let therefore $(Y_n)_{n\geq 0}$ be a sequence of two-dimensional random vectors satisfying the recurrence

$$Y_n \stackrel{d}{=} A_1(n)\, Y_{I_n} + A_2(n)\, Y'_{J_n}, \qquad n \geq n_0, \tag{5.7}$$

where $\left(A_2(n)\,, A_2(n)\,, I^{(n)}\right), (Y_n)$ and (Y'_n) are independent, $A_1(n)$ and $A_2(n)$ are random 2×2 matrices, (I_n, J_n) is a vector of random cardinalities $I_n, J_n \in \{0, \ldots, n\}$ and (Y'_n) is identically distributed as (Y_n). Furthermore $n_0 \geq 1$, and in the case $2 < s \leq 3$, we assume that $\mathrm{Cov}(Y_n)$ is positive definite for $n \geq n_1 \geq n_0$.

We have to control the moments of order up to s, so we normalize Y_n by setting

$$X_n := C_n^{-1/2}\left(Y_n - M_n\right), \qquad n \geq 0, \tag{5.8}$$

where $M_n := \mathbb{E}[Y_n]$. In the case $1 < s \leq 2$, C_n is a positive definite square matrix, whereas in the case $2 < s \leq 3$, we set

$$C_n := \begin{cases} \mathrm{Id}_2 & 0 \leq n < n_1, \\ \mathrm{Cov}(Y_n) & n \geq n_1. \end{cases}$$

The normalized quantities (X_n) then satisfy the modified recurrence

$$X_n \stackrel{d}{=} A_1^{(n)} X_{I_n} + A_2^{(n)} X'_{J_n} + b'(n)\,, \qquad n \geq n_1, \tag{5.9}$$

with

$$\begin{aligned} A_1^{(n)} &:= C_n^{-1/2} A_1(n)\, C_{I_n}^{1/2} \\ A_2^{(n)} &:= C_n^{-1/2} A_2(n)\, C_{J_n}^{1/2} \\ b'(n) &:= C_n^{-1/2}\left(A_1(n)\, M_{I_n} + A_2(n)\, M_{J_n} - M_n\right). \end{aligned}$$

Theorem 5.1 (Neininger and Rüschendorf [26, Thm 4.1]). *Let X_n be given as in* (5.8) *and be s-integrable, $1 < s \leq 3$. We assume that*

$$\left(A_1^{(n)}, A_2^{(n)}, b'(n)\right) \xrightarrow{\ell_s} (A_1, A_2, b)\,, \tag{5.10}$$

$$\mathbb{E}\Big[\|A_1\|_{\mathrm{op}}^s\Big] + \mathbb{E}\Big[\|A_2\|_{\mathrm{op}}^s\Big] < 1, \tag{5.11}$$

$$\mathbb{E}\big[\mathbb{1}_{\{I_n<l\}\cup\{I_n=n\}}\big] + \mathbb{E}\big[\mathbb{1}_{\{J_n<l\}\cup\{J_n=n\}}\big] \to 0 \tag{5.12}$$

for all $l \in \mathbb{N}$. Then X_n converges weakly to a limit X with distribution characterized by the fixed-point equation

$$X \stackrel{d}{=} A_1 X + A_2 X' + b', \tag{5.13}$$

where (A_1, A_2, b'), X, and X' are independent and X' is distributed as X. In the case $1 < s \leq 2$, the distribution $\mathcal{L}(X)$ is given as the unique solution of this equation in $\mathcal{M}_s^{\mathbb{R}^d}(0)$, whereas in the case $2 < s \leq 3$, the fixed-point is unique in $\mathcal{M}_s^{\mathbb{R}^d}(0, \mathrm{Id}_2)$.

5.2.3 Application

The approach described in the last section can now be used for the example of the urn with random replacements of section 4.2. We will see that the requirements on the coefficients, especially condition (5.11), are too restrictive and only satisfied for a fraction of the possible values of α and β. For the normal limit case, a further disadvantage of this approach, when compared to the method used in the preceding chapters, is that one needs the covariance of the random vector. We will not discuss how to get this, but merely investigate the consequences of a reasonable assumption.

The non-normal limit case. Assume that $\lambda := \alpha + \beta - 1 > 1/2$ and denote by $\mu(n)$ the vector consisting of the expectations $\mu_{\mathrm{b}}(n)$ and $\mu_{\mathrm{w}}(n)$, as defined in (4.33) and given explicitely in Lemma 4.3. For $n \to \infty$, this implies that

$$\mu(n) = c'n + d'n^\lambda + \mathrm{o}\left(n^\lambda\right),$$

with constants as in (4.36),

$$c' = \begin{pmatrix} c'_{\mathrm{b}} \\ c'_{\mathrm{w}} \end{pmatrix} = \frac{1-\beta}{1-\lambda}\begin{pmatrix} 1 \\ 1 \end{pmatrix}, \quad d' = \begin{pmatrix} d'_{\mathrm{b}} \\ d'_{\mathrm{w}} \end{pmatrix} = \frac{1}{(1-\lambda)\,\Gamma(\lambda+1)}\begin{pmatrix} 1-\alpha \\ \beta-1 \end{pmatrix}. \tag{5.14}$$

We use the normalization $X_0 := (0,0)^t$ and

$$X_n := \frac{1}{n^\lambda}\big(B_n - \mu(n)\big) \quad , n \geq 1. \tag{5.15}$$

From recursion (5.1), we obtain for the normalized quantity, and $n \geq 1$, the recursive equation

$$X_n \stackrel{d}{=} \left(\frac{I_n}{n}\right)^\lambda X_{I_n} + \left(\frac{J_n}{n}\right)^\lambda \begin{pmatrix} F_\alpha & 1-F_\alpha \\ 1-F_\beta & F_\beta \end{pmatrix} X'_{J_n} + b'(n)\,, \tag{5.16}$$

where

$$b'(n) = \left(\left(\frac{I_n}{n}\right)^\lambda - 1\right) d' + \left(\frac{J_n}{n}\right)^\lambda \begin{pmatrix} F_\alpha & 1-F_\alpha \\ 1-F_\beta & F_\beta \end{pmatrix} d' + \mathrm{o}(1)$$

with conditions on independence and identical distributions analogous to (5.1). This suggests for the limit X of X_n the fixed-point equation

$$X \stackrel{d}{=} U^\lambda X + (1-U)^\lambda \begin{pmatrix} F_\alpha & 1-F_\alpha \\ 1-F_\beta & F_\beta \end{pmatrix} X' + b', \tag{5.17}$$

with

$$b' = \left(U^\lambda - 1\right) d' + (1-U)^\lambda \begin{pmatrix} F_\alpha & 1-F_\alpha \\ 1-F_\beta & F_\beta \end{pmatrix} d', \tag{5.18}$$

where X, X', U, and (F_α, F_β) are independent and $X \stackrel{d}{=} X'$.

When investigating the requirements of Theorem 5.1 with $s = 2$, conditions (5.10) and (5.12) are clearly satisfied in our case. For condition (5.11), we have to check that

$$\xi_2 := \mathbb{E}\Big[U^{2\lambda}\Big] + \mathbb{E}\Big[(1-U)^{2\lambda}\Big]\,\mathbb{E}\left[\left\|\begin{pmatrix} F_\alpha & 1-F_\alpha \\ 1-F_\beta & F_\beta \end{pmatrix}\right\|_{\mathrm{op}}^2\right] < 1. \tag{5.19}$$

The replacement matrix has a quite simple form which enables us to easily determine the operator norm, getting for any $s > 0$:

$$\mathbb{E}\left[\left\|\begin{pmatrix} F_\alpha & 1-F_\alpha \\ 1-F_\beta & F_\beta \end{pmatrix}\right\|_{\mathrm{op}}^s\right] = \mathbb{E}\Big[(1 + |F_\alpha - F_\beta|)^{s/2}\Big] \tag{5.20}$$

We minimize this by choosing an appropriate joint distribution of F_α and F_β: Let V be uniformly distributed on the unit interval, independent of everything else, and set $F_\alpha = \mathbb{1}_{\{V \le \alpha\}}$ and $F_\beta = \mathbb{1}_{\{V \le \beta\}}$. Then for any $s > 0$,

$$\mathbb{E}\Big[(1 + |F_\alpha - F_\beta|)^{s/2}\Big] = 1 + |\alpha - \beta| \left(2^{s/2} - 1\right), \tag{5.21}$$

so condition (5.19) is satisfied for $s = 2$ if

$$\xi_2 = \frac{2 + |\alpha - \beta|}{2\lambda + 1} < 1, \tag{5.22}$$

which, for $\lambda > 1/2$, is always true if $\alpha = \beta$ but does not hold for all possible values of α and β with $\alpha + \beta - 1 > 1/2$. In figure 5.1, the values for which this condition is satisfied are marked grey. Note that the condition in the ℓ_2-variant of the theorem (cf. Neininger [25, Thm 4.1]), although weaker in general, leads to exactly the same condition for this example.

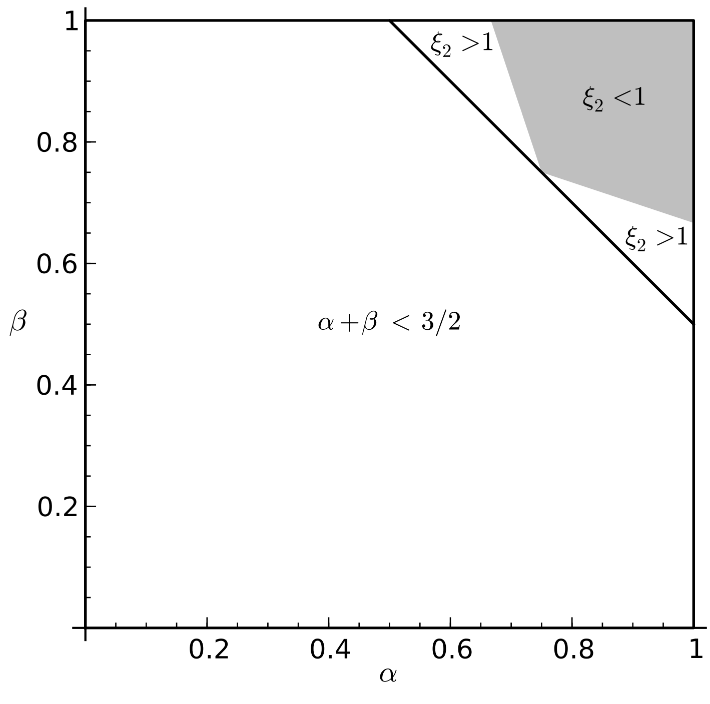


Figure 5.1: Valid combinations of α and β for which condition (5.11) of the multivariate limit theorem of Neininger and Rüschendorf [26, Thm 4.1] (or the corresponding condition in Neininger [25, Thm 4.1]) is satisfied for $s = 2$

Remark. One idea to get convergence for the remaining combinations of α and β might be to increase s. As in the normal limit case below, this would require to find the covariance matrix of B_n. Furthermore, even for $s = 3$, condition (5.11) requires

$$\xi_3 = \frac{2 + |\alpha - \beta| \left(2^{3/2} - 1\right)}{3\lambda + 1} < 1, \tag{5.23}$$

which is in fact a weaker condition than (5.22), but again is not satisfied for all possible combinations of α and β, as indicated in figure 5.2.

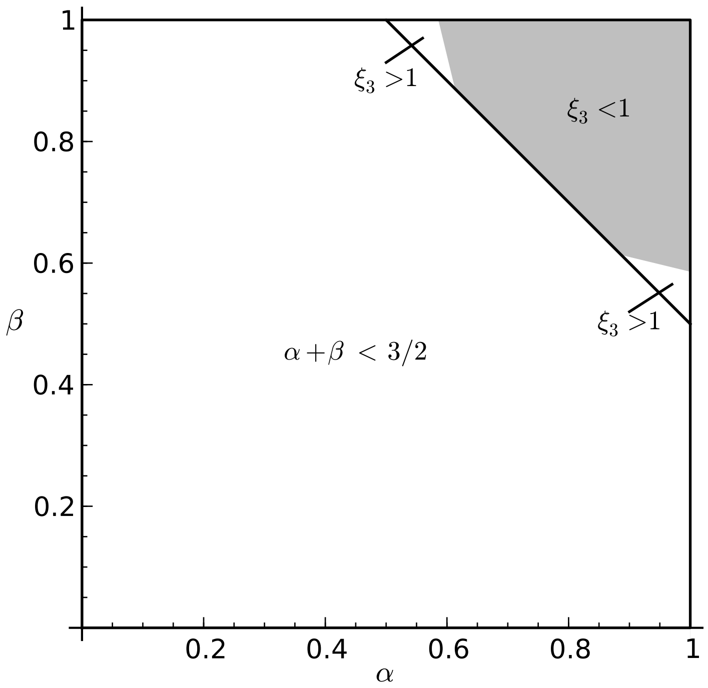


Figure 5.2: Valid combinations of α and β for which condition (5.11) of the multivariate limit theorem of Neininger and Rüschendorf [26, Thm 4.1] is satisfied for $s = 3$

The normal limit case. We only outline the case $\lambda \coloneqq \alpha + \beta - 1 < 1/2$. The expansions in Lemma 4.3 imply, for $n \to \infty$, that

$$\mu(n) = c'n + \mathrm{o}\big(\sqrt{n}\big)\,,$$

with a constant c' derived from (4.36)

$$c' = \begin{pmatrix} c'_{\mathrm{b}} \\ c'_{\mathrm{w}} \end{pmatrix} = \frac{1-\beta}{1-\lambda} \begin{pmatrix} 1 \\ 1 \end{pmatrix}. \tag{5.24}$$

Analogously to the normal limit cases in the examples in chapter 4, we have to work with the Zolotarev metric with index $s > 2$ on $\mathcal{M}_s^{\mathbb{R}^2}(\mu, C)$. This requires finding the covariance matrix of B_n, which also depends on the joint distribution of F_α and F_β. Note that in chapter 4 we only needed the variance of the components.

We will not give an expression for the covariance here but only assume that a linear expression can be found at least for the case with joint distribution of F_α and F_β as used in the non-normal limit case above. More precisely, we will assume that a symmetric, positive definite 2×2 matrix f' can be found, such that asymptotically

$$\mathrm{Cov}(B_n) = f'n + \mathrm{o}(n)\,. \tag{5.25}$$

This implies that $\mathrm{Cov}(B_n)$ is positive definite for all $n \geq n_1$ for appropriate n_1. We now define

$$C_n \coloneqq \begin{cases} \mathrm{Id}_2 & 0 \leq n < n_1 \\ \mathrm{Cov}(B_n) & n \geq n_1 \end{cases} \tag{5.26}$$

and use this for the normalization. We set $X_0 \coloneqq (0,0)^t$ and for $n \geq 1$

$$X_n \coloneqq C_n^{-1/2}\,(B_n - \mu)\,. \tag{5.27}$$

According to (5.9), we get for the scaled quantity a recursive equation of the form

$$X_n \stackrel{d}{=} A_1^{(n)} X_{I_n} + A_2^{(n)} X'_{J_n} + b'(n)\,, \qquad n \geq n_1, \tag{5.28}$$

where

$$\begin{aligned} A_1^{(n)} &= \left(\frac{I_n}{n}\right)^{1/2} + g_1(n) \\ A_2^{(n)} &= \left(\frac{J_n}{n}\right)^{1/2} (f')^{-1/2} \begin{pmatrix} F_\alpha & 1 - F_\alpha \\ 1 - F_\beta & F_\beta \end{pmatrix} (f')^{1/2} + g_2(n)\,. \end{aligned}$$

The matrix f' is symmetric and positive definite and therefore can be diagonalized, i.e. there is an orthogonal matrix S and a positive definite diagonal matrix D such that $f' = SDS^t$. For powers of f', we have $(f')^\alpha = SD^\alpha S^t$, where D^α is just the power α applied to each diagonal element of D. Substituting this in the expression for $A_2^{(n)}$ and using that for any square matrix A multiplication by a diagonal matrix is commutative, $DA = AD$, we get the recursive equation

$$X_n \stackrel{d}{=} \left(\frac{I_n}{n}\right)^{1/2} X_{I_n} + \left(\frac{J_n}{n}\right)^{1/2} \begin{pmatrix} F_\alpha & 1-F_\alpha \\ 1-F_\beta & F_\beta \end{pmatrix} X'_{J_n} + e'(n)\,, \tag{5.29}$$

with conditions on independence and identical distributions analogous to (5.1). Similarly to section 4.2, $e'(n)$ vanishes in the limit. This suggests for the limit X of X_n the fixed-point equation

$$X \stackrel{d}{=} \sqrt{U}X + \sqrt{1-U} \begin{pmatrix} F_\alpha & 1-F_\alpha \\ 1-F_\beta & F_\beta \end{pmatrix} X', \tag{5.30}$$

where X, X', U, and (F_α, F_β) are independent and X' is distributed as X.

To show convergence using Theorem 5.1, this time using $s = 3$, it remains to check that

$$\xi := \mathbb{E}\Big[U^{3/2}\Big] + \mathbb{E}\Big[(1-U)^{3/2}\Big]\, \mathbb{E}\left[\left\| \begin{pmatrix} F_\alpha & 1-F_\alpha \\ 1-F_\beta & F_\beta \end{pmatrix} \right\|_{\mathrm{op}}^3\right] < 1. \tag{5.31}$$

With joint distribution of F_α and F_β as in the non-normal case and using (5.20) and (5.21) for $s = 3$, we get for ξ the expression

$$\xi = \frac{2}{5}\Big(2 + |\alpha - \beta|\left(2^{3/2} - 1\right)\Big). \tag{5.32}$$

This implies that, similar to the non-normal limit case, the condition is satisfied only for a fraction of the possible combinations of α and β satisfying $\alpha + \beta \leq 3/2$. In figure 5.3, the respective combinations are marked grey.

5.3 Extension of the convergence theorem

The Zolotarev metric, as proposed by Zolotarev [37, 38], can be defined for distributions not only on Euclidean space but more generally on Banach spaces. It turns out that endowing $\mathbb{R}^d$ with a p-norm for $p \in [1, \infty]$ allows us to extend Theorem 5.1 in such a way that it covers our example of an urn with random replacement.

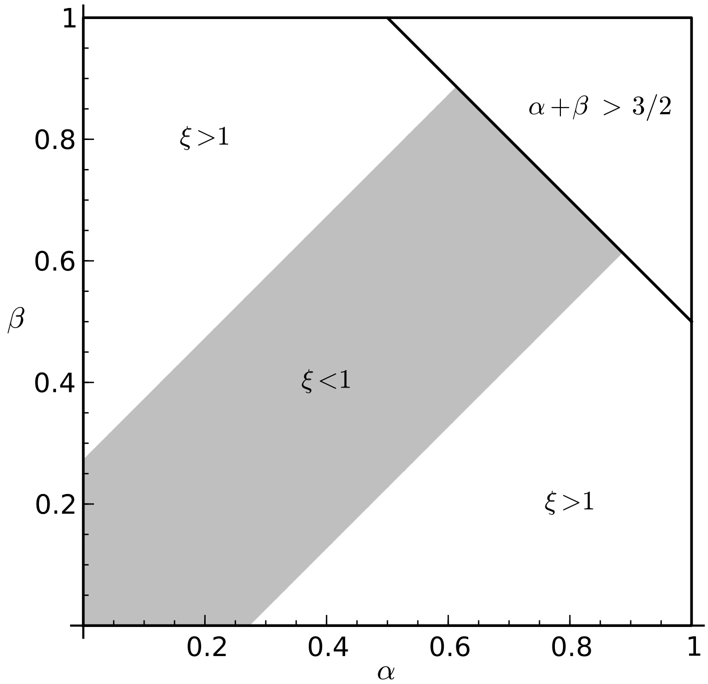


Figure 5.3: Valid combinations of α and β for which condition (5.11) of the multivariate limit theorem of Neininger and Rüschendorf [26, Thm 4.1] is satisfied for $s = 3$

Zolotarev metric. We first define our version of the Zolotarev metric which is used subsequently. For $p \in [1, \infty]$ and $x = (x_1 \ldots, x_d) \in \mathbb{R}^d$ the p-norm is denoted by

$$\|x\|_p := \left(\sum_{j=1}^{d} |x_j|^p \right)^{1/p}, \quad 1 \leq p < \infty, \text{ and}$$
$$\|x\|_\infty := \max_{1 \leq j \leq d} |x_j|$$

We use the concept of m times (Fréchet-)differentiable functions from $\mathbb{R}^d$ to $\mathbb{R}$:

Definition 5.2 (Fréchet-differentiability)**.** Let $(V, \|\cdot\|_V)$ and $(W, \|\cdot\|_W)$ be Banach spaces and $U \subset V$ an open subset of V. A function $f : U \to W$ is called Fréchet-differentiable in $x \in U$, if there exists a bounded linear operator $A_x : V \to W$ such that

$$\lim_{h \to 0} \frac{\|f(x+h) - f(h) - A_x(h)\|_W}{\|h\|_V} = 0.$$

Further, f is called differentiable in U, if it is differentiable in every $x \in U$.

In general, this definition depends on the norms used on the Banach spaces. However, the p-norms on $\mathbb{R}^d$ are equivalent: for any $p, q \in [1, \infty]$ there exist constants $c_{p,q} > 0$ such that

$$\|x\|_p \leq c_{p,q} \|x\|_q \text{ for all } x \in \mathbb{R}^d. \tag{5.33}$$

So if a function $f : \mathbb{R}^d \to \mathbb{R}$ is Fréchet-differentiable with respect to some p-norm or some other norm on $\mathbb{R}^d$, this is true for any $p \geq 1$ and any other norm and the resulting derivatives are equal. Therefore the space $C^m(\mathbb{R}^d, \mathbb{R})$ of all real-valued functions on $\mathbb{R}^d$ which are m times continuously differentiable is well defined and does not depend on our choice of p.

The mth derivative of f is a function from $\mathbb{R}^d$ into the space $\mathscr{L}\left(\left(\mathbb{R}^d\right)^m, \mathbb{R}\right)$ of multilinear mappings $\left(\mathbb{R}^d\right)^m \to \mathbb{R}$. So Hölder-continuity of order α with Hölder-constant 1 of the mth derivative translates into the condition

$$\left\| f^{(m)}(x) - f^{(m)}(y) \right\|_{\mathrm{op}(p)} \leq \|x - y\|_p^\alpha \tag{5.34}$$

for all $x, y \in \mathbb{R}^d$, where $\|\cdot\|_{\mathrm{op}(p)}$ denotes the operator norm with respect to the p-norm, defined for any multilinear mapping $L \in \mathscr{L}\left(\left(\mathbb{R}^d\right)^m, \mathbb{R}\right)$ by

$$\|L\|_{\mathrm{op}(p)} := \sup_{\|h_1\|_p, \dots, \|h_m\|_p \leq 1} |L(h_1, \dots, h_m)| .$$

Therefore, we now define the family of test functions $\mathcal{F}_{s,p}$, depending on $p \in [1, \infty]$, as

$$\mathcal{F}_{s,p} := \left\{ f \in C^m\left(\mathbb{R}^d, \mathbb{R}\right) \;\middle|\; \forall\, x, y \in \mathbb{R}^d : \left\| f^{(m)}(x) - f^{(m)}(y) \right\|_{\mathrm{op}(p)} \leq \|x - y\|_p^\alpha \right\}.$$

Using this, we can define the Zolotarev metric $\zeta_{s,p}$ as before, using $\mathcal{F}_{s,p}$ as family of test functions:

$$\zeta_{s,p}(X, Y) := \zeta_{s,p}\big(\mathcal{L}(X), \mathcal{L}(Y)\big) := \sup_{f \in \mathcal{F}_{s,p}} \left| \mathbb{E}[f(X) - f(Y)] \right|, \tag{5.35}$$

where $s = m + \alpha$ with $0 < \alpha \leq 1$, $m \in \mathbb{N}_0$ and $p \in [1, \infty]$.

It follows from general results on Zolotarev metrics on Banach spaces, see [37, 38], that for any $p \in [1, \infty]$ and for $0 < s \leq 1$, we have that $\zeta_{s,p}(X, Y) < \infty$ for all $\mathcal{L}(X), \mathcal{L}(Y) \in \mathcal{M}_s^{\mathbb{R}^d}$ and $\left(\mathcal{M}_s^{\mathbb{R}^d}, \zeta_{s,p}\right)$ is a metric space. For $1 < s \leq 2$, we have that $\zeta_{s,p}(X,Y) < \infty$ for all $\mathcal{L}(X), \mathcal{L}(Y) \in \mathcal{M}_s^{\mathbb{R}^d}(\mu)$ and any $\mu \in \mathbb{R}^d$ and that $\left(\mathcal{M}_s^{\mathbb{R}^d}(\mu), \zeta_{s,p}\right)$ is a metric space. Finally, for $2 < s \leq 3$, we have that $\zeta_{s,p}(X,Y) < \infty$ for any $\mathcal{L}(X), \mathcal{L}(Y) \in \mathcal{M}_s^{\mathbb{R}^d}(\mu, C)$ for any $\mu \in \mathbb{R}^d$ and symmetric, positive semidefinite matrix C and also that $\left(\mathcal{M}_s^{\mathbb{R}^d}(\mu, C), \zeta_{s,p}\right)$ is a metric space.

We have the following estimates between the metrics $\zeta_{s,p}$ for different p:

Lemma 5.3. *Let $\zeta_{s,p}$ be defined as in* (5.35) *and $p, q \in [1, \infty]$. Then for constants $c_{p,q}$ with* (5.33) *we have*

$$\zeta_{s,p} \leq c_{p,q}^s \zeta_{s,q}.$$

Proof. For the operator norm of any multilinear mapping L we get, using constants with (5.33),

$$\begin{aligned}
\|L\|_{\mathrm{op}(q)} &= \sup_{\|h_1\|_q, \ldots, \|h_m\|_q \leq 1} \left|L(h_1, \ldots, h_m)\right| \\
&= \sup_{\|c_{p,q}h_1\|_q, \ldots, \|c_{p,q}h_m\|_q \leq 1} \left|L(c_{p,q}h_1, \ldots, c_{p,q}h_m)\right| \\
&= c_{p,q}^m \sup_{c_{p,q}\|h_1\|_q, \ldots, c_{p,q}\|h_m\|_q \leq 1} \left|L(h_1, \ldots, h_m)\right| \\
&\leq c_{p,q}^m \sup_{\|h_1\|_p, \ldots, \|h_m\|_p \leq 1} \left|L(h_1, \ldots, h_m)\right| \\
&= c_{p,q}^m \|L\|_{\mathrm{op}(p)}.
\end{aligned}$$

For any function $f \in \mathcal{F}_{s,p}$ we can conclude that for any $q \geq 1$

$$\begin{aligned}
\left\|f^{(m)}(x) - f^{(m)}(y)\right\|_{\mathrm{op}(q)} &\leq c_{p,q}^m \left\|f^{(m)}(x) - f^{(m)}(y)\right\|_{\mathrm{op}(p)} \\
&\leq c_{p,q}^m \|x - y\|_p^\alpha \\
&\leq c_{p,q}^m \cdot c_{p,q}^\alpha \|x - y\|_q^\alpha \\
&= c_{p,q}^s \|x - y\|_q^\alpha.
\end{aligned}$$

This implies in particular for any $f \in \mathcal{F}_{s,p}$ that $c_{p,q}^{-s} f \in \mathcal{F}_{s,q}$, or vice versa that if $c_{p,q}^{s} f \in \mathcal{F}_{s,p}$ then $f \in \mathcal{F}_{s,q}$. Using this relation, we get for any $\mathbb{R}^d$-valued random variables X and Y, that

$$\begin{aligned}
\zeta_{s,p}(X,Y) &= \sup_{f \in \mathcal{F}_{s,p}} \left| \mathbb{E}[f(X) - f(Y)] \right| \\
&= \sup_{c_{p,q}^s g \in \mathcal{F}_{s,p}} \left| \mathbb{E}\left[c_{p,q}^s g(X) - c_{p,q}^s g(Y)\right] \right| \\
&= c_{p,q}^s \cdot \sup_{c_{p,q}^s g \in \mathcal{F}_{s,p}} \left| \mathbb{E}[g(X) - g(Y)] \right| \\
&\le c_{p,q}^s \cdot \sup_{g \in \mathcal{F}_{s,q}} \left| \mathbb{E}[g(X) - g(Y)] \right| \\
&= c_{p,q}^s \, \zeta_{s,q}(X,Y)\,.
\end{aligned}$$

□

The properties of ζ_s cited in section 3.1 also hold for $\zeta_{s,p}$ for any $p \in [1,\infty]$. In particular, the corresponding spaces are complete and convergence with respect to $\zeta_{s,p}$ implies weak convergence on $\mathbb{R}^d$, see also [10, Theorem 5.1].

Careful inspection of the proof of the multivariate limit theorem of Neininger and Rüschendorf [26, Thm 4.1] shows that the theorem still holds for $\zeta_{s,p}$ using the operator norm with respect to the p-norm.

Theorem 5.4. *Theorem 5.1 also holds when condition* (5.11) *is replaced by*

$$\mathbb{E}\left[\|A_1\|_{\mathrm{op}(p)}^s\right] + \mathbb{E}\left[\|A_2\|_{\mathrm{op}(p)}^s\right] < 1, \tag{5.36}$$

for any $p \in [1,\infty]$.

Remark. Using the supremum norm ($p = \infty$) significantly simplifies the computation for condition (5.36). The operator norm with respect to the supremum norm is just the maximum of the absolute row sums of the matrix, i.e. the sums of the absolute values of the entries in each row.

5.4 Convergence

Using theorem 5.4 we can now prove weak convergence for the non-normal limit case. In the normal limit case, condition (5.36) is also satisfied. However, to finish the proof an expression for the covariance matrix of B_n would be needed.

The non-normal limit case. Recall that $\lambda := \alpha+\beta-1 > 1/2$. Starting as in section 5.2.3, we use the normalization (5.15) to get the recursive equation (5.16) which in turn suggest for the limit X the fixed-point equation

$$X \stackrel{d}{=} U^\lambda X + (1-U)^\lambda \begin{pmatrix} F_\alpha & 1-F_\alpha \\ 1-F_\beta & F_\beta \end{pmatrix} X' + b', \tag{5.37}$$

with

$$b' = \left(U^\lambda - 1\right) d' + (1-U)^\lambda \begin{pmatrix} F_\alpha & 1-F_\alpha \\ 1-F_\beta & F_\beta \end{pmatrix} d' \tag{5.38}$$

where X, X', U, and (F_α, F_β) are independent and $X \stackrel{d}{=} X'$.

To use theorem 5.4 with $s=2$ it remains to check that

$$\mathbb{E}\left[U^{2\lambda}\right] + \mathbb{E}\left[(1-U)^{2\lambda}\right] \mathbb{E}\left[\left\|\begin{pmatrix} F_\alpha & 1-F_\alpha \\ 1-F_\beta & F_\beta \end{pmatrix}\right\|_{\mathrm{op}(p)}^2\right] < 1. \tag{5.39}$$

Using the operator norm with respect to the p-norm, we can replace (5.20) for $s>0$ by

$$\mathbb{E}\left[\left\|\begin{pmatrix} F_\alpha & 1-F_\alpha \\ 1-F_\beta & F_\beta \end{pmatrix}\right\|_{\mathrm{op}(p)}^s\right] = \begin{cases} \mathbb{E}\left[(1+|F_\alpha - F_\beta|)^{s/p}\right] & 1 \le p < \infty \\ 1 & p = \infty, \end{cases} \tag{5.40}$$

so if we use the supremum norm on $\mathbb{R}^d$, (5.39) is clearly satisfied. At the same time, we could couple F_α and F_β as before, and get for any $s>0$,

$$\mathbb{E}\left[(1+|F_\alpha - F_\beta|)^{s/p}\right] = 1 + |\alpha-\beta|\left(2^{s/p}-1\right). \tag{5.41}$$

Using this, condition (5.39) is satisfied if

$$\frac{2+|\alpha-\beta|\left(2^{2/p}-1\right)}{2\lambda+1} < 1, \tag{5.42}$$

which, for $\lambda > 1/2$, is always true if $\alpha = \beta$ and otherwise can be ensured by choosing

$$p > \frac{2\log 2}{\log\left(1 + \frac{2\lambda-1}{|\alpha-\beta|}\right)}$$

which is always possible as long as $\lambda > 1/2$.

The normal limit case. We only sketch the case $\lambda := \alpha + \beta - 1 < 1/2$. Starting as in section 5.2, we use the normalization (5.27) leading to the recursive equation (5.29) and suggesting for the limit X the fixed-point equation (5.30).

To show convergence using theorem 5.4, this time using $s = 3$, it remains to check that

$$\mathbb{E}\left[U^{s/2}\right] + \mathbb{E}\left[(1-U)^{s/2}\right]\mathbb{E}\left[\left\|\begin{pmatrix} F_\alpha & 1-F_\alpha \\ 1-F_\beta & F_\beta \end{pmatrix}\right\|_{\mathrm{op}(p)}^{s}\right] < 1. \tag{5.43}$$

Using (5.40) it is easy to see that this is satisfied when using the supremum norm on $\mathbb{R}^d$. For $1 \leq p < \infty$ this condition is satisfied for $s = 3$ if and only if

$$\frac{2}{5}\left(2 + |\alpha-\beta|\left(2^{3/p} - 1\right)\right) < 1, \tag{5.44}$$

which can be ensured for any combination of α and β satisfying $\alpha + \beta \leq 3/2$ by choosing $p \geq 6$.

Comparing the results in this chapter with the approach in the first chapters using a system of recurrence relations, the main disadvantage of this approach is that in the normal limit case the covariance matrix of the random vector is needed. Furthermore, when thinking of future applications of the approach, it is not clear if it is possible to find a suitable p in all cases where it might be possible to prove convergence using the approach using a system of recurrence equations. On the other hand, if the conditions of Theorem 5.4 are satisfied, proving convergence is possible in a straightforward way.

Bibliography

[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(6):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(3):394–405.

[4] Bai, Z.-D. and Hu, F. (1999). Asymptotic theorems for urn models with nonhemogeneous 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. (2012a). Limit distributions for multitype branching processes of m-ary search trees. *Ann. IHP*, to appear.

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

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

[10] Drmota, M., Janson, S., and Neininger, R. (2008). A functional limit theorem for the profile of search trees. *Ann. Appl. Probab.*, 18(1):288–333.

[11] Eggenberger, F. and Pólya, G. (1923). Über die Statistik verketteter Vorgänge. *Z. angew. Math. Mech.*, 3:279–289.

[12] Fill, J. A. and Kapur, N. (2004). The space requirement of m-ary search trees: Distributional asymptotics for $m \geq 27$. In *Proceedings of the 7th Iranian Statistical Conference, 2004.*

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

[14] Janson, S. (1983). Limit theorems for certain branching random walks on compact groups and homogeneous spaces. *Ann. Probab.*, 11(4):909–930.

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

[16] Janson, S. (2006). Limit theorems for triangular urn schemes. *Probab. Theory Related Fields*, 134(3):417–452.

[17] Janson, S. and Neininger, R. (2008). The size of random fragmentation trees. *Probab. Theory Related Fields*, 142:399–442.

[18] Johnson, N. and Kotz, S. (1977). *Urn models and their application. An approach to modern discrete probability theory.* Wiley Series in Probability and Mathematical Statistics. John Wiley & Sons, New York-London-Sydney.

[19] Kaijser, S. and Janson, S. (2012). Higher moments of banach space valued random variables. arXiv:1208.4272.

[20] Knape, M. and Neininger, R. (2013). Pólya urns via the contraction method. arXiv: 1301.3404.

[21] Kotz, S., Mahmoud, H., and Robert, P. (2000). On generalized Pólya urn models. *Statist. Probab. Lett.*, 49:163–173.

[22] 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 *Proceedings of ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 877–886.

[23] Mahmoud, H. M. (2009). *Pólya urn models.* Texts in Statistical Science Series. CRC Press, Boca Raton, FL.

[24] Matthews, P. C. and Rosenberger, W. F. (1997). Variance in randomized play-the-winner clinical trials. *Statist. Probab. Lett.*, 35:233–240.

[25] Neininger, R. (2001). On a multivariate contraction method for random recursive structures with applications to quicksort. *Random Structures and Algorithms*, 19(3-4):498–524.

[26] Neininger, R. and Rüschendorf, L. (2004). A general limit theorem for recursive algorithms and combinatorial structures. *Ann. Appl. Probab.*, 14(1):378–418.

[27] Neininger, R. and Rüschendorf, L. (2006). A survey of multivariate aspects of the contraction method. *Discrete Math. Theor. Comput. Sci.*, 8:31–56.

[28] Neininger, R. and Sulzbach, H. (2012). On a functional contraction method. arXiv:1202.1370.

[29] Pouyanne, N. (2005). Classification of large Pólya-Eggenberger urns with regard to their asymptotics. *DMTCS proc.*, AD:275–286.

[30] Pouyanne, N. (2008). An algebraic approach to pólya processes. *Henri Poincaré Probab. Stat.*, 44(2):193–323.

[31] Rachev, S. and Rüschendorf, L. (1995). Probability metrics and recursive algorithms. probability metrics and recursive algorithms. probability metrics and recursive algorithms. *Adv. in Appl. Probab.*, 27:770–799.

[32] Rösler, U. (1991). A limit theorem for "quicksort". *RAIRO Inform. Théor. Appl.*, 25:85–100.

[33] Smythe, R. (1996). Central limit theorems for urn models. *Stochastic Process. Appl.*, 65:115–137.

[34] Smythe, R. T. and Rosenberger, W. F. (1995). Play-the-winner designs, generalized Pólya urns, and markov branching processes. In *Adaptive Designs*, volume 25 of *IMS Lecture Notes - Monograph Series*. IMS.

[35] Wei, L. J. and Durham, S. (1978). The randomized play-the-winner rule in medical trials. *J. Amer. Statist. Assoc.*, 73(364):840–843.

[36] 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(409):156–162.

[37] Zolotarev, V. M. (1976). Approximation of the distributions of sums of independent random variables with values in infinite-dimensional spaces. *Theory Probab. Appl.*, 21:721–737.

[38] Zolotarev, V. M. (1977). Ideal metrics in the problem of approximating distributions of sums of independent random variables. *Theory Probab. Appl.*, 22(3):433–449.

Zusammenfassung

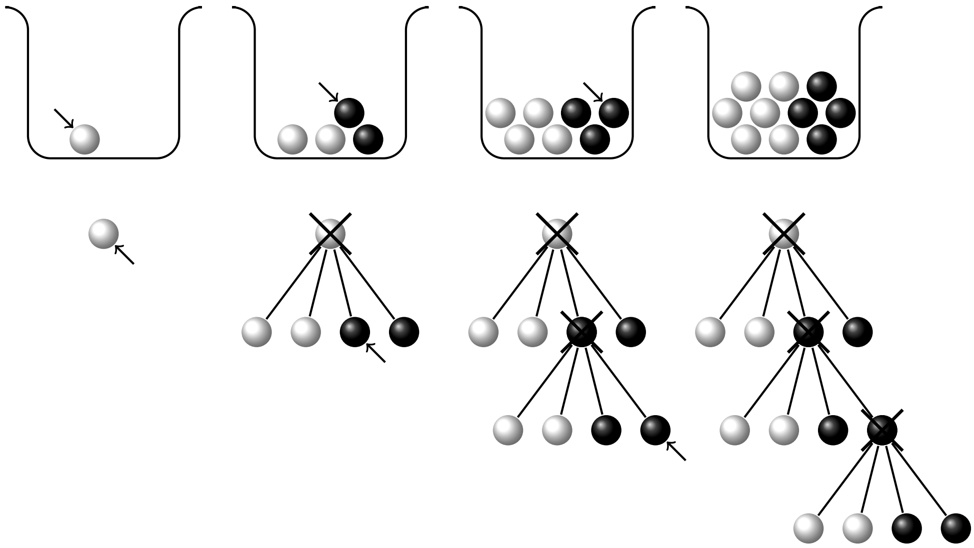
Pólya-Urnenmodelle haben vielfältige Einsatzmöglichkeiten und wurden, abhängig von der jeweiligen Fragestellung, bereits mit verschiedensten Methoden untersucht. So kann zum Beispiel die Entwicklung der Anzahl der Blätter in einem Binärsuchbaum durch eine Urne modelliert und das asymptotische Verhalten aus allgemeineren Resultaten über Urnenmodelle abgelesen werden. In dieser Arbeit wird der umgekehrte Weg beschritten und die Entwicklung der Urne mit der Entwicklung eines geeigneten Baumes assoziiert, dessen rekursive Struktur dann zur Analyse genutzt wird. Dieser Zugang bietet insbesondere den Vorteil, dass mit elementaren Mitteln alle drei grundsätzlich verschiedenen Möglichkeiten für das Langzeitverhalten, die in diesem Modell möglich sind, herausgearbeitet werden können.

Das hier untersuchte Urnenmodell besteht dabei aus einer Urne, die zu Beginn eine Kugel enthält, die eine von m Farben hat. Die Urne entwickelt sich schrittweise, indem jeweils eine Kugel rein zufällig aus der Urne gezogen wird. Abhängig von der Farbe der gezogenen Kugel wird diese zusammen mit einer festen Anzahl $K-1$ an weiteren Kugeln, möglicherweise in verschiedenen Farben, zurückgelegt. Die Farben der nachgelegten Kugeln werden durch eine Nachlegematrix beschrieben, in der die Zeilen jeweils der Farbe der gezogenen Kugel entsprechen und die Spalten den Farben der nachgelegten Kugeln. Die Einträge der Matrix können von einer zusätzlichen Zufallsquelle abhängen, ihre Verteilung muss jedoch zu Beginn festgelegt werden. Die Zeilensummen sind stets deterministisch gleich $K-1$.

Die zugrundeliegende Idee für die Analyse ist nun, die Entwicklung der Urne durch die Entwicklung eines vollständigen K-nären Baumes zu beschreiben und die dadurch sichtbar werdende rekursive Struktur zu nutzen, um mit Hilfe der Kontraktionsmethode Aussagen über das Langzeitverhalten zu machen.

Im assoziierten Baum entspricht jedes Blatt einer Kugel in der Urne. Wird eine Kugel aus der Urne gezogen, so wird das entsprechende Blatt im Baum zum inneren

Knoten. Die inneren Knoten im Baum entsprechen nicht (mehr) bestimmten Kugeln in der Urne, ihre Farbe ist jedoch für die Zerlegung noch wichtig. Für die zurück- bzw. nachgelegten Kugeln werden K Blätter an diesen Knoten angefügt, deren Farbe durch die Nachlegematrix bestimmt wird. Das folgende Bild zeigt die Enwicklung einer Pólya-Urne mit zwei Farben, weiß und schwarz, und Nachlegematrix $\left[\begin{smallmatrix}1 & 2\\ 2 & 1\end{smallmatrix}\right]$ sowie den zugeordneten Baum. Die Pfeile zeigen jeweils an, welche Kugel gezogen wird.



Je nach Farbe der Startkugel gibt es verschiedene Typen dieser assoziierten Bäume. Um eine rekursive Darstellung zu erhalten, zerlegt man den Baum an der Wurzel in die K Teilbäume für die direkten Kinder. Bedingt auf die jeweilige Anzahl der Blätter sind die Teilbäume unabhängige assoziierte Bäume der entsprechenden Größe, deren Typ durch die Farbe der jeweiligen Wurzel festgelegt ist. Die Anzahl der Blätter in den jeweiligen Teilbäumen kann durch eine Pólya-Urne modelliert werden, bei der die Nachlegematrix gerade das $(K-1)$-fache der Einheitsmatrix ist. Für diese ist bekannt, dass die Anteile der jeweiligen Farben fast sicher gegen einen Dirichlet-verteilten Zufallsvektor konvergieren.

Für die Verteilungen der assoziierten Bäume kann nun ein System von Rekursionsgleichungen aufgestellt werde. Wenn man dabei mehr als zwei Farben zulässt, bietet es sich an, diese mit Zahlen $1, \ldots, m$ zu bezeichnen. Die Nachlegematrix hat dann die Form $R = (a_{ij})_{1 \leq i,j \leq m}$ mit

$$a_{ij} \in \begin{cases} \mathbb{N}_0, & \text{for } i \neq j, \\ \mathbb{N}_0 \cup \{-1\}, & \text{for } i = j, \end{cases} \qquad \text{und} \qquad \sum_{j=1}^{m} a_{ij} =: K - 1 \geq 1 \text{ for } i = 1, \ldots, m.$$

Sei nun $B_n^{[j]}$ die Anzahl der Blätter mit Farbe 1 nach n Schritten in einem Baum, dessen Wurzel die Farbe j hat, und $I^{(n)} = (I_1^{(n)}, \ldots, I_K^{(n)})$ der Vektor der Anzahlen der Blätter in den jeweiligen Teilbäumen. Um die Rekursionsgleichungen zu formulieren, führen wir ferner die Intervalle

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

ein, mit der Konvention, dass $[x, y] = \emptyset$ falls $x > y$. Damit ergibt sich für den assoziierten Baum mit Wurzel in Farbe j die Rekursionsgleichung

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

wobei für jedes $j \in \{1, \ldots, m\}$ die Familie

$$\left\{ \left(B_k^{[i],(r)} \right)_{0 \leq k < n} \,\middle|\, r \in J_{ij}, i \in \{1, \ldots, m\} \right\} \cup \left\{ I^{(n)} \right\}$$

unabhängig ist, $B_k^{[i],(r)}$ dieselbe Verteilung hat wie $B_k^{[i]}$ für alle $i \in \{1, \ldots, m\}$, $0 \leq k < n$ und $r \in J_{ij}$. Der Vektor $I^{(n)}$ ist dabei asymptotisch Dirichlet-verteilt.

Dieses Modell kann auch auf zufällige Nachlegematrizen erweitert werden, d.h. in jedem Schritt werden die Einträge gemäß einer gegebenen Verteilung neu festgelegt. Ein Beispiel hierfür wird in Abschnitt 4.2 diskutiert.

Abhängig von der Nachlegematrix können drei typische Fälle für das asymptotische Verhalten auftreten, wobei wir nur solche Urnen untersuchen, für die die Anzahl der schwarzen Kugeln asymptotisch linear wächst:

(a) Die Erwartungswerte haben für $n \to \infty$ die Form

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

mit einer Konstanten $c_\mu > 0$, die nicht von der Anfangsfarbe j abhängt und reellwertigen Konstanten d_j und einem Exponenten $1/2 < \lambda < 1$. Die Varianzen sind hier typischerweise von der Größenordnung $n^{2\lambda}$ und nach geeigneter Skalierung bietet sich für den Grenzwert das System von Fixpunktgleichungen

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

an, wobei $X^{[i],(r)}$ und $(D_1, \dots, D_K)$ unabhängig sind, $X^{[i],(r)}$ dieselbe Verteilung wie $X^{[i]}$ hat, $(D_1, \dots, D_K)$ Dirichlet-verteilt ist und die $b^{[j]}$ Funktionen von $(D_1, \dots, D_K)$ sind. Die Verteilungen der $X^{[j]}$ sind in diesem Fall keine Normalverteilungen.

(b) Die Erwartungswerte haben für $n \to \infty$ die Form

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

mit einer Konstante $c_\mu > 0$, die nicht von der Startfarbe j abhängt. Die Varianzen sind hier linear und nach geeigneter Skalierung bietet sich für den Grenzwert das System von Fixpunktgleichungen

$$X^{[j]} \stackrel{d}{=} \sum_{i=1}^{m} \sum_{r \in J_{ij}} \sqrt{D_r} X^{[i],(r)}, \quad j = 1, \dots, m \tag{2}$$

an, mit Bedingungen wie bei (1). Unter geeigneten Annahmen an die Momente hat dieses System eine eindeutige Lösung: alle $X^{[j]}$ sind standardnormalverteilt (Satz 3.2). Dasselbe gilt für den Fall, dass

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

bei dem jedoch die Größenordnung der Varianzen $n \log^\delta(n)$ ist, mit $\delta > 0$.

(c) Für Urnen mit mehr als zwei Farben ist es auch möglich, dass die Erwartungswerte für $n \to \infty$ die Form

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

haben, mit einer Konstante $c_\mu > 0$, die unabhängig von der Startfarbe ist, einem Exponenten $1/2 < \lambda < 1$ und Konstanten $\kappa_j \in \mathbb{C}$ und $\mu \in \mathbb{R}$. Obwohl die Oszillation von Erwartungswert und Varianz typischerweise nicht durch geeignete Skalierung beseitigt werden kann und dadurch eine Konvergenz gegen eine feste Grenzverteilung verhindert wird, kann durch Übergang auf die Komplexe Zahlenebene der Grenzwert durch das System von Fixpunktgleichungen

$$X^{[j]} \stackrel{d}{=} \sum_{i=1}^{m} \sum_{r \in J_{ij}} D_r^{\omega} X^{[i],(r)}, \quad j = 1, \ldots, m \tag{3}$$

beschrieben werden, wobei wiederum die Bedingungen aus (1) gelten und zudem $\omega := \lambda + \mathrm{i}\mu$. Unter geeigneten Annahmen an die Momente hat dieses System eine eindeutige Lösung unter den Verteilungen auf $\mathbb{C}$ (Satz 3.3).

Um die Existenz und Eindeutigkeit der Lösungen zu zeigen, wird das d-fache kartesische Produkt des Raumes der s-integrierbaren reellwertigen Zufallsvariablen (mit der Borel'schen σ-Algebra) $\mathcal{M}_s^{\mathbb{R}}$ (bzw. entsprechend $\mathcal{M}_s^{\mathbb{C}}$ für komplexwertige Zufallsvariablen) mit einer Produkt-Version der Zolotarevmetrik (ζ_2 bzw. ζ_3) oder der minimalen L_2-Metrik (ℓ_2) ausgestattet. Die Teilräume (mit fixiertem erstem bzw. zweitem Moment) werden so gewählt, dass die Metrik endlich und der metrisierte Raum vollständig ist und daher mit dem Banach'schen Fixpunktsatz geschlossen werden kann.

Dieselben Metrischen Räume werden auch verwendet, um für drei Beispiele die Konvergenz gegen die Grenzverteilung zu zeigen (Kapitel 4). Bei der Urne mit zwei Farben treten nur die Fälle (a) und (b) auf. Dies gilt sowohl für das erste Beispiel mit deterministischer Nachlegematrix, als auch für das zweite, bei dem die Nachlegematrix zufällig ist. Im dritten Beispiel sind mehr als zwei Farben möglich und auch Fall (c) kann eintreten. Die Beweise sind möglichst allgemein gehalten, so dass sie auch auf andere Beispiele übertragen werden können.

Im letzten Kapitel wird noch eine etwas andere Herangehensweise untersucht, die näher an bisherigen Anwendungen der Kontraktionsmethode liegt. Dazu wird anstelle eines Systems von Rekursionsgleichungen nur eine einzige Rekursion für einen mehrdimensionalen Zufallsvektor formuliert. Für solche mehrdimensionalen Rekursionsgleichungen gibt es einen Satz von Neininger and Rüschendorf [26, Thm. 4.1], der Bedingungen an die Koeffizienten aufstellt und dann Existenz und Eindeutigkeit des Grenzwertes und (schwache) Konvergenz liefert.

Eine direkte Anwendung dieses Satzes auf das zweite Beispiel scheitert daran, dass die Voraussetzungen des Satzes nicht für jede zulässige Kombination der Parameter erfüllt sind. Durch eine Veränderung der zugrundeliegenden Metrik lässt sich jedoch zeigen, dass der Satz dennoch auf diesen Fall angewendet werden kann. Allerdings bleiben die Voraussetzungen stärker, als für die in den ersten Kapiteln beschriebene Herangehensweise. Insbesondere muss im Fall (b) nicht nur die Varianz der einzelnen Zufallsvariablen, sondern die Kovarianzmatrix kontrolliert werden, was in Anwendungen ein deutliches Hindernis darstellen kann. Andererseits ist die vorgestellte Erweiterung des Konvergenzsatzes von Neininger und Rüschendorf auch für andere Anwendungen interessant, da die veränderten Bedingungen an die Koeffizienten nicht nur schwächer, sondern im Allgemeinen auch einfacher nachzuweisen sind.