Rates of convergence for balanced irreducible two-color Pólya urns

Andrea Kuntschik & Ralph Neininger *

Abstract

For balanced Pólya urns with two colors the (normalized) number of balls of each color satisfies a limit law with two possible regimes: with weak convergence towards the normal distribution and almost sure convergence towards distributions that can be characterized by moments or by recursive distributional equations. We bound the rate of convergence in these limit theorems for such irreducible urn schemes. The bounds are sufficiently tight to confirm a conjecture of S. Janson for a subclass of these urns.

MSC2010: 60C05, 60F05, 60J05, 68Q25.

Keywords: Pólya urn, Pólya-Eggenberger urn, weak convergence, limit law, probability metric, rate of convergence, recursive distributional equation, contraction method.

1 Introduction and results

We consider a Pólya urn with balls of two possible colors, black and white. The dynamics of the urn are determined by the initial configuration and a replacement matrix. For simplicity we assume that the urn process starts with one ball. The replacement matrix is denoted by

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

Time evolves in discrete steps. In each step, one ball is drawn uniformly at random from the urn. If the ball withdrawn is black, it is placed back into the urn together with a black and b white balls. If it is white, it is placed back together with c black and d white balls. An entry -1 in the replacement matrix indicates that the drawn ball is not replaced back into the urn (but still balls of the other color could be added). The steps are iterated independently.

Throughout this extended abstract, we assume that the urn is *balanced*, i.e., that $a+b = c+d =: K-1 \geq 1$, and *irreducible*, i.e., that $bc \neq 0$.

We denote by B_n the number of black balls after n steps, hence $B_0 \in \{0, 1\}$ depending on whether we start with a black or a white ball. Note that the number of white balls after n steps is $(K-1)n+(1-B_0)-B_n$. The asymptotic behavior of B_n for the balanced, irreducible 2×2 Pólya urn, as described above, is fully known with respect to limit laws, see [2, 9]. Depending on the ratio λ of the two eigenvalues of R,

$$\lambda := \frac{a-c}{a+b},$$

we have two regimes: If $\lambda \leq \frac{1}{2}$ and $\lambda \neq 0$, then

$$Z_n := \frac{B_n - \mathbb{E}[B_n]}{\sqrt{\mathrm{Var}(B_n)}} \xrightarrow{d} \mathcal{N}(0,1), \tag{1.1}$$

where $\xrightarrow{d}$ denotes convergence in distribution and $\mathcal{N}(0,1)$ the standard normal distribution. All asymptotic statements in this extended abstract are as $n \to \infty$. (In the case $\lambda = 0$ the evolution of the urn is deterministic.) If $\lambda > \frac{1}{2}$, then we have almost surely that

$$X_n := \frac{B_n - \mathbb{E}[B_n]}{n^{\lambda}} \longrightarrow X_R, \tag{1.2}$$

with a limit distribution $\mathcal{L}(X_R)$ depending on the replacement matrix R, see (2.6)–(2.7).

The subject of this extended abstract is to provide bounds on the rates of convergence in the limit laws (1.1) and (1.2). We use the following three metrics: The Kolmogorov–Smirnov metric ϱ, the minimal L_p metrics ℓ_p, and the Zolotarev metric ζ_3 defined as follows: We denote by F_V the distribution function of a random variable V. Then, the Kolmogorov–Smirnov distance (uniform distance) is given by

$$\varrho(V,W) := \varrho(\mathcal{L}(V), \mathcal{L}(W)) := \sup_{x \in \mathbb{R}} |F_V(x) - F_W(x)|.$$

The minimal L_p-metric is given by

(1.3)
$$\ell_p(V,W) := \ell_p(\mathcal{L}(V), \mathcal{L}(W)) := \inf\{\|V' - W'\|_p : \mathcal{L}(V) = \mathcal{L}(V'), \mathcal{L}(W) = \mathcal{L}(W')\},$$

*Institute of Mathematics,
Goethe University, Box 187
60629 Frankfurt am Main, Germany
E-mail: {kuntschik,neiningr}@math.uni-frankfurt.de

for all $1 \le p < \infty$ and random variables V, W with $\|V\|_p, \|W\|_p < \infty$. Note that the infimum in (1.3) is over all joint distributions $\mathcal{L}(V', W')$ with the given marginals $\mathcal{L}(V)$ and $\mathcal{L}(W)$. Finally, the Zolotarev metric ζ_s with $s = 3$ is given by

$$\zeta_3(V,W) := \zeta_3(\mathcal{L}(V), \mathcal{L}(W)) := \sup_{f \in \mathcal{F}_3} |\mathbb{E}[f(V) - f(W)]|,$$

where

$$\mathcal{F}_3 := \{f \in C^2(\mathbb{R}, \mathbb{R}) : |f^{(2)}(x) - f^{(2)}(y)| \le |x - y|\}$$

denotes the space of twice continuously differentiable functions from $\mathbb{R}$ to $\mathbb{R}$ such that the second derivative is Lipschitz continuous with Lipschitz-constant 1. We have the following results:

Theorem 1.1. *If $\lambda > \frac{1}{2}$, then we have for all $\varepsilon > 0$ and all $1 \le p < \infty$ that*

$$\ell_p(X_n, X_R) = \mathrm{O}\left(n^{-\lambda + \frac{1}{2} + \varepsilon}\right),$$
$$\varrho(X_n, X_R) = \mathrm{O}\left(n^{-\lambda + \frac{1}{2} + \varepsilon}\right).$$

Theorem 1.2. *If $\lambda \le \frac{1}{2}$, then we have for all $\varepsilon > 0$ that*

$$\zeta_3(Z_n, \mathcal{N}(0,1)) = \begin{cases} \mathrm{O}\left(n^{-\frac{1}{2}+\varepsilon}\right), & \text{for } \lambda \le \frac{1}{3},\ \lambda \neq 0, \\ \mathrm{O}\left(n^{3(\lambda - \frac{1}{2})}\right), & \text{for } \frac{1}{3} < \lambda < \frac{1}{2}, \\ \mathrm{O}\left((\log n)^{-\frac{1}{2}}\right), & \text{for } \lambda = \frac{1}{2}. \end{cases}$$

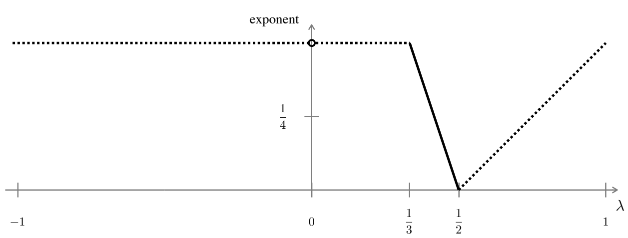

Figure 1: Shown are the exponents in the decay of the bounds on the rates of convergence in Theorems 1.1 and 1.2. For $\lambda = \frac{1}{2}$ there is no polynomial bound, see Theorem 1.2. The dashed line indicates that our exponents are (arbitrarily close) below the values shown.

We have no lower bounds on the rates of convergence in Theorems 1.1 and 1.2; hence we have no information on the optimality of these bounds. We weakly conjecture that the ε in Theorems 1.1 and 1.2 can be dropped in all bounds and that the orders $n^{-\lambda+\frac{1}{2}}$ and $n^{-\frac{1}{2}}$ are optimal, respectively. About the case $\lambda = \frac{1}{2}$ in Theorem 1.2 the authors are divided about whether $(\log n)^{-\frac{1}{2}}$ or $(\log n)^{-\frac{3}{2}}$ should be the correct rate in ζ_3.

The rate $n^{3(\lambda - \frac{1}{2})}$ for $\frac{1}{3} < \lambda < \frac{1}{2}$ in Theorem 1.2 was mentioned in Janson [9, Remark 4.7]. Janson did not specify the metric and said it was "tempting to conjecture" this rate. Moreover, he noted there that his methods (in [9]) give no information on the rate of convergence.

For a discussion of relations between rates of convergence in the metrics ϱ, ℓ_p and ζ_3 and implications of such rates see [5, 15]. A contraction proof for asymptotic normality as given in [12], being the starting point for our derivation of bounds on the rate of convergence, requires a Zolotarev metric ζ_s with $s \in (2, 3]$. However, we have not checked whether the bounds given in Theorem 1.2 extend to any ζ_s with $2 < s < 3$.

Note that the rates of convergence of Theorems 1.1 and 1.2 can be extended to arbitrary initial configurations of the urn, see the discussions in [3, p. 933–934] and [12, p. 1165].

For general references on the analysis of urn models we refer to the monographs of Johnson and Kotz [11] and Mahmoud [14] and the references and comments on the literature in the papers of Janson [9], Flajolet et al. [6] and Pouyanne [18]. Note that for a few specific replacement matrices rates of convergence have been bounded before, see, e.g., [8, 6, 7, 17].

Although it is hidden in this extended abstract, one essential ingredient of our approach is a good control on the asymptotics of the variance of B_n. For the two-color urns covered in the present note we draw back to an explicit formula, see [19, 2]. In view of the recent preprint [10] we hope to be able to extend our results to irreducible, balanced urn models with an arbitrary number of colors.

2 Sketch of the proofs of the Theorems 1.1 and 1.2

We first recall a system of recursive distributional equations satisfied by B_n in Section 2.1. In Section 2.2, a bound of a rate of convergence for an auxiliary urn process is presented. The proofs of Theorems 1.1 and 1.2 are then sketched in Sections 2.3 and 2.4 respectively.

2.1 Recursive distributional equations The bounds in Theorems 1.1 and 1.2 are proved by making estimates from Knape and Neininger in [12] explicit. Their approach was based on a tree-structure underlying the evolution of the urn that yields a system of recursive distributional equations. We sketch how these equations are derived since our approach totally rests on these equations. For a detailed account see Section

2 of [12].

Initially, there is one ball in the urn. It is encoded by a tree that consists of one node, a leaf. In each step, a ball is drawn and returned to the urn together with $K-1$ new balls. In terms of the tree, this corresponds to picking a leaf and adding K children to this leaf. Additionally, all leaves are labelled by their colors, e.g., $a+1$ black and b white leaves will be added to the root in the first step if the root is black and c black and $d+1$ white leaves if the root is white, respectively, see Figure 2.

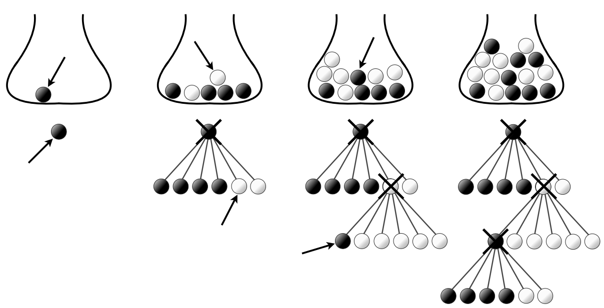

Figure 2: Evolution of an urn with replacement matrix $\left[\begin{smallmatrix}3 & 2\\ 1 & 4\end{smallmatrix}\right]$ together with its associated tree: Three steps of the urn process are shown when beginning with one black ball. Below the urn the respective associated tree is shown. The arrow indicates which ball is drawn. The leaves of the associated tree correspond to the balls in the urn (for the sake of clarity, the internal nodes are marked with a cross).

This tree evolves simultaneously to the urn. We call the tree so constructed the associated tree. We deal with two kinds of associated trees: One tree emerging from a black root, another emerging from a white root. To distinguish between these two, we call the former b-associated tree, the latter w-associated tree. By construction, the number of black balls in the urn and the number of black leaves of the associated tree coincide. We will count them as the sum of the numbers of the black leaves within the K subtrees of the root. By $I^{(n)} := \left(I_1^{(n)}, \dots, I_K^{(n)}\right)$ we denote a random vector of integers whose components $I_r^{(n)}$ describe how often a leaf from the r-th subtree was picked within the first n steps of the evolution of the urn. Note that we have $\sum_{r=1}^K I_r^{(n)} = n-1$ and that the r-th subtree has $I_r^{(n)}(K-1)+1$ leaves.

Finally, we observe that the K subtrees of the root conditioned on $I^{(n)}$ behave independently and are distributed as b- and w-associated trees, respectively, possessing the respective numbers of leaves.

Thus, we obtain the following recursive representation of the number of black balls after n steps subject to the color of the initial ball with $B_0^b := 1$ and $B_0^w := 0$ and for $n \geq 1$, see [12],

$$B_n^b \overset{d}{=} \sum_{r=1}^{a+1} B_{I_r^{(n)}}^{b,(r)} + \sum_{r=a+2}^{K} B_{I_r^{(n)}}^{w,(r)}, \tag{2.4}$$

$$B_n^w \overset{d}{=} \sum_{r=1}^{c} B_{I_r^{(n)}}^{b,(r)} + \sum_{r=c+1}^{K} B_{I_r^{(n)}}^{w,(r)} \tag{2.5}$$

with $B_j^{b,(r)} \overset{d}{=} B_j^b$, $B_j^{w,(r)} \overset{d}{=} B_j^w$ for $r = 1, \dots, K$ and $0 \leq j \leq n$ and $\left(B_j^{b,(1)}\right)_{0\leq j\leq n}, \dots, \left(B_j^{b,(K)}\right)_{0\leq j\leq n}$, $\left(B_j^{w,(1)}\right)_{0\leq j\leq n}, \dots, \left(B_j^{b,(K)}\right)_{0\leq j\leq n}$, $I^{(n)}$ being independent.

2.2 Asymptotics of $I^{(n)}$ It follows from Athreya [1] that $\frac{1}{n}I^{(n)}$ converges almost surely to a Dirichlet$(\frac{1}{K}, \dots, \frac{1}{K})$ distributed vector $D = (D_1, \dots, D_K)$. To bound the rates of convergence in Theorems 1.1 and 1.2, we need a bound on the rate of convergence in the latter limit first.

LEMMA 2.1. *For all $p \geq 2$ there is $E_p > 0$ such that for $\alpha \in (0,1]$, $r = 1, \dots, K$ and $n \in \mathbb{N}$ it holds*

$$\left\| \left(\frac{I_r^{(n)}}{n}\right)^{\alpha} - D_r^{\alpha} \right\|_p \leq E_p \cdot n^{-\frac{\alpha}{2}}.$$

Proof. We give a sketch: We interpret $(I_1^{(n)}(K-1)+1, \dots, I_K^{(n)}(K-1)+1)_{n\in\mathbb{N}_0}$ as the evolution of the numbers of balls of types $1, \dots, K$ in an urn with initially one ball of each of the types $1, \dots, K$ and with a diagonal matrix as replacement matrix with $K-1$ on each of its diagonal entries, cf. Lemma 2.1 in [12]. This urn process is covered by [1, Corollary 1] from which $\frac{1}{n}I^{(n)} \to D$ almost surely follows. We denote by $Y_n := (Y_{n,1}, \dots, Y_{n,K})$, $n \geq 1$, the outcome of the n-th draw, i.e., $Y_{n,j} = 1$ and $Y_{n,i} = 0, i \neq j$, if a ball of type j was drawn, $j \in \{1, \dots, K\}$. Obviously, we have $I^{(n)} = \sum_{i=1}^n Y_i$. An explicit calculation shows that the sequence $(Y_n)_{n\geq 1}$ is exchangeable. Hence, we may apply de Finetti's Theorem to obtain independence of the Y_n conditional on D. For independent summands we use the Marcinkiewicz-Zygmund inequality [4, p.386]:

> *For every $p \geq 1$ there exist constants $\kappa_p, \tau_p > 0$ such that for any sequence $(W_n)_{n\geq 1}$ of independent random variables with mean 0 we have*

$$\kappa_p \left\| \left(\sum_{j=1}^n W_j^2\right)^{\frac{1}{2}} \right\|_p \leq \left\| \sum_{j=1}^n W_j \right\|_p \leq \tau_p \left\| \left(\sum_{j=1}^n W_j^2\right)^{\frac{1}{2}} \right\|_p .$$

Combining the conditional independence with the Marcinkiewicz-Zygmund inequality implies for all $r \in \{1,\dots,K\}$ that

$$\left\|\frac{1}{n}I_r^{(n)} - D_r\right\|_p \quad \le \quad \frac{M_p}{\sqrt{n}}$$

with a suitable constant $M_p > 0$. The claim of the Lemma follows by Jensen's inequality.

Recently, Svante Janson informed us that he has tight bounds on the distance between $\frac{I_r^{(n)}}{n}$ and D_r of the order $\frac{1}{n}$ for any ℓ_p, $1 \le p \le \infty$. In particular, this shows that our $n^{-\frac{\alpha}{2}}$ in Lemma 2.1 can be improved to $n^{-\alpha}$.

2.3 Sketch of the proof of Theorem 1.1 We set $\mu_b(n) := \mathbb{E}[B_n^b], \mu_w(n) := \mathbb{E}[B_n^w]$, $X_0 = Y_0 = 0$ and for $n \ge 1$

$$X_n := \frac{B_n^b - \mu_b(n)}{n^\lambda}, \quad Y_n := \frac{B_n^w - \mu_w(n)}{n^\lambda}.$$

Then the system (2.4)–(2.5) of distributional recurrences for B_n^b and B_n^w turns into

$$X_n \stackrel{d}{=} \sum_{r=1}^{a+1}\left(\frac{I_r^{(n)}}{n}\right)^\lambda X^{(r)}_{I_r^{(n)}} + \sum_{r=a+2}^{K}\left(\frac{I_r^{(n)}}{n}\right)^\lambda Y^{(r)}_{I_r^{(n)}} + b_b(n),$$
$$Y_n \stackrel{d}{=} \sum_{r=1}^{c}\left(\frac{I_r^{(n)}}{n}\right)^\lambda X^{(r)}_{I_r^{(n)}} + \sum_{r=c+1}^{K}\left(\frac{I_r^{(n)}}{n}\right)^\lambda Y^{(r)}_{I_r^{(n)}} + b_w(n)$$

with toll terms $b_b(n)$ and $b_w(n)$ given by

$$\begin{aligned} b_b(n) &= \frac{\sum_{r=1}^{a+1}\mu_b\left(I_r^{(n)}\right) + \sum_{r=a+2}^{K}\mu_w\left(I_r^{(n)}\right) - \mu_b(n)}{n^\lambda} \\ &= d_b\left(-1 + \sum_{r=1}^{a+1}\left(\frac{I_r^{(n)}}{n}\right)^\lambda\right) + d_w\sum_{r=a+2}^{K}\left(\frac{I_r^{(n)}}{n}\right)^\lambda \\ &\quad + O\left(n^{-\lambda}\right), \\ b_w(n) &= \frac{\sum_{r=1}^{c}\mu_b\left(I_r^{(n)}\right) + \sum_{r=c+1}^{K}\mu_w\left(I_r^{(n)}\right) - \mu_b(n)}{n^\lambda} \\ &= d_b\sum_{r=1}^{c}\left(\frac{I_r^{(n)}}{n}\right)^\lambda + d_w\left(-1 + \sum_{r=c+1}^{K}\left(\frac{I_r^{(n)}}{n}\right)^\lambda\right) \\ &\quad + O\left(n^{-\lambda}\right), \end{aligned}$$

with appropriate constants $d_b, d_w \in \mathbb{R}$ and conditions on (conditional) independence and distributions similar to (2.4)–(2.5). The limit behavior of these sequences is covered by Janson [9]. In the context of the contraction method, limit distributions are identified by the system

$$X \stackrel{d}{=} \sum_{r=1}^{a+1} D_r^\lambda X^{(r)} + \sum_{r=a+2}^{K} D_r^\lambda Y^{(r)} + b_b, \tag{2.6}$$

$$Y \stackrel{d}{=} \sum_{r=1}^{c} D_r^\lambda X^{(r)} + \sum_{r=c+1}^{K} D_r^\lambda Y^{(r)} + b_w \tag{2.7}$$

with toll terms

$$b_b = d_b\left(-1 + \sum_{r=1}^{a+1} D_r^\lambda\right) + d_w\sum_{r=a+2}^{K} D_r^\lambda,$$
$$b_w = d_b\sum_{r=1}^{c} D_r^\lambda + d_w\left(-1 + \sum_{r=c+1}^{K} D_r^\lambda\right)$$

with independent copies $X^{(r)}$ of X, $Y^{(r)}$ of Y, $r = 1,\dots,K$, $D = (D_1,\dots,D_K)$ as above, where $X^{(1)},\dots,X^{(K)}, Y^{(1)},\dots,Y^{(K)}$ and D are independent, see [12, 3]. Among all pairs of centered probability distributions there is a unique fixed point subsequently denoted by $(\mathcal{L}(\Lambda_b), \mathcal{L}(\Lambda_w))$ to the system (2.6)–(2.7) giving the limit distributions of the $(X_n)_{n\ge1}$ and $(Y_n)_{n\ge1}$ respectively, see [12].

To bound rates of convergences, we denote

$$\Delta_b(n) := \ell_2(X_n, \Lambda_b), \quad \Delta_w(n) := \ell_2(Y_n, \Lambda_w),$$
$$\Delta(n) := \Delta_b(n) \vee \Delta_w(n).$$

By use of optimal couplings and estimates being standard in the context of the contraction method, then

$$\begin{aligned} &\Delta^2(n) \\ &\le \sum_{r=1}^{K}\left(\mathbb{E}\left[\left(\frac{I_r^{(n)}}{n}\right)^{2\lambda}\Delta^2\left(I_r^{(n)}\right)\right]\right. \\ &\quad + L^2\left\|\left(\frac{I_r^{(n)}}{n}\right)^\lambda - D_r^\lambda\right\|_2^2 \\ &\quad \left. + 2L\mathbb{E}\left[\left|\left(\frac{I_r^{(n)}}{n}\right)^\lambda\left(\left(\frac{I_r^{(n)}}{n}\right)^\lambda - D_r^\lambda\right)\right|\Delta\left(I_r^{(n)}\right)\right]\right) \\ &\quad + \max\left\{\|b_b(n) - b_b\|_2^2, \|b_w(n) - b_w\|_2^2\right\} \end{aligned} \tag{2.8}$$

with $L := \max\{\|\Lambda_b\|_2, \|\Lambda_w\|_2\}$.

From Lemma 2.1 with $p = 2$ and $\alpha = \lambda$ we obtain

$$\|b_b(n) - b_b\|_2 \vee \|b_w(n) - b_w\|_2 = O\left(n^{-\frac{\lambda}{2}}\right). \tag{2.9}$$

First, from (2.8) with (2.9) we obtain the ℓ_p bound in Theorem 1.1 for $p = 2$ by induction.

Second, this ℓ_2 bound is extended to $p > 2$ by use of a direct extension of [5, Lemma 3.2]:

LEMMA 2.2. *Let $V_1, \ldots, V_{K+1}$, $K \geq 2$ be independent random variables and $p \geq 2$ integer. Then,*

$$\mathbb{E}\left[\left|\sum_{i=1}^{K+1} V_i\right|^p\right] \leq \sum_{i=1}^{K} \mathbb{E}[|V_i|^p] + \left(\sum_{i=1}^{K} \|V_i\|_{p-1} + \|V_{K+1}\|_p\right)^p.$$

We start (with appropriate optimal couplings) with

$$\begin{aligned} &\ell_p\left(X_n, \Lambda_b\right) \\ \leq\ & \left\| \left\{ \sum_{r=1}^{a+1} \left(\frac{I_r^{(n)}}{n}\right)^\lambda X^{(r)}_{I_r^{(n)}} - D_r^\lambda X^{(r)} \right\} \right. \\ &+ \left\{ \sum_{r=a+2}^{K} \left(\frac{I_r^{(n)}}{n}\right)^\lambda Y^{(r)}_{I_r^{(n)}} - D_r^\lambda Y^{(r)} \right\} \\ &\left. + b_b\left(n\right) - b_b \right\|_p \\ =:\ & \left\| \sum_{r=1}^{a+1} V_r + \sum_{r=a+2}^{K} V_r + V_{K+1} \right\|_p \end{aligned}$$

The summands $V_1, \ldots, V_{K+1}$ are conditioned on the random vector $\left(D, I^{(n)}\right)$ independent; hence, we can apply Lemma 2.2. Then we do an induction proof over $p \geq 2$ where for each step $p \to p+1$ we do an induction as in the case $p = 2$ over n. This leads to the ℓ_p bounds in Theorem 1.1 for $1 \leq p < \infty$.

To obtain the bound in the Kolmogorov–Smirnov metric ϱ stated in Theorem 1.1, we use

LEMMA 2.3. [5, Lemma 5.1] *Assume, $\mu, \nu \in \mathcal{M}$ such that μ is absolutely continuous with a bounded density function f. Let $M := \sup_{x\in\mathbb{R}} |f(x)|$ and $1 \leq p < \infty$. Then,*

$$\varrho\left(\mu, \nu\right) \leq \left(p+1\right)^{\frac{1}{p+1}} \left(M \ell_p\left(\mu, \nu\right)\right)^{\frac{p}{p+1}}.$$

To apply Lemma 2.3, we only need to know that Λ_b, Λ_w have bounded densities. This was shown in [13]. The bound on ϱ in Theorem 1.1 follows.

2.4 Sketch of the proof of Theorem 1.2 We set $\sigma_b(n) := \sqrt{\mathrm{Var}(B_n^b)}, \sigma_w(n) := \sqrt{\mathrm{Var}(B_n^w)}$, $Z_0 = \Upsilon_0 = Z_1 = \Upsilon_1 = 0$, and, for $n \geq 2$,

$$Z_n := \frac{B_n^b - \mu_b(n)}{\sigma_b(n)}, \quad \Upsilon_n := \frac{B_n^w - \mu_w(n)}{\sigma_w(n)}.$$

Then the system (2.4)–(2.5) of distributional recurrences for B_n^b and B_n^w turns into

$$\begin{aligned} Z_n &\stackrel{d}{=} \sum_{r=1}^{a+1} \frac{\sigma_b\left(I_r^{(n)}\right)}{\sigma_b(n)} Z^{(r)}_{I_r^{(n)}} + \sum_{r=a+2}^{K} \frac{\sigma_w\left(I_r^{(n)}\right)}{\sigma_b(n)} \Upsilon^{(r)}_{I_r^{(n)}} + t_b\left(I^{(n)}\right), \\ \Upsilon_n &\stackrel{d}{=} \sum_{r=1}^{c} \frac{\sigma_b\left(I_r^{(n)}\right)}{\sigma_w(n)} Z^{(r)}_{I_r^{(n)}} + \sum_{r=c+1}^{K} \frac{\sigma_w\left(I_r^{(n)}\right)}{\sigma_w(n)} \Upsilon^{(r)}_{I_r^{(n)}} + t_w\left(I^{(n)}\right) \end{aligned}$$

with

$$\begin{aligned} t_b\left(I^{(n)}\right) &= \frac{\sum_{r=1}^{a+1} \mu_b\left(I_r^{(n)}\right) + \sum_{r=a+2}^{K} \mu_w\left(I_r^{(n)}\right) - \mu_b(n)}{\sigma_b(n)}, \\ t_w\left(I^{(n)}\right) &= \frac{\sum_{r=1}^{c} \mu_b\left(I_r^{(n)}\right) + \sum_{r=c+1}^{K} \mu_w\left(I_r^{(n)}\right) - \mu_w(n)}{\sigma_w(n)}, \end{aligned}$$

and conditions on (conditional) independence and identical distributions as in (2.4)–(2.5). For $\lambda \leq \frac{1}{2}$ (and $\lambda \neq 0$) expansions of $\mu_b, \mu_w, \sigma_b, \sigma_w$ imply a limit system of the form

$$Z \stackrel{d}{=} \sum_{r=1}^{a+1} \sqrt{D_r} Z^{(r)} + \sum_{r=a+2}^{K} \sqrt{D_r} \Upsilon^{(r)}, \tag{2.10}$$

$$\Upsilon \stackrel{d}{=} \sum_{r=1}^{c} \sqrt{D_r} Z^{(r)} + \sum_{r=c+1}^{K} \sqrt{D_r} \Upsilon^{(r)}, \tag{2.11}$$

with independent copies $Z^{(r)}$ of Z and independent copies $\Upsilon^{(r)}$ of Υ, $r = 1, \ldots, K$, all independent of D. From [12, Theorem 5.2] we have that the (under appropriate conditions on moments) unique fixed-point solving (2.10) and (2.11) is $(\mathcal{N}(0,1), \mathcal{N}(0,1))$. Now, we set

$$\Delta'(n) := \zeta_3(Z_n, \mathcal{N}(0,1)) \vee \zeta_3(\Upsilon_n, \mathcal{N}(0,1)).$$

We use a typical construction in the framework of the contraction method with ζ_s metrics, so called accompanying sequences:

$$\begin{aligned} Q_n^b &:= \sum_{r=1}^{a+1} \frac{\sigma_b\left(I_r^{(n)}\right)}{\sigma_b(n)} N_r + \sum_{r=a+2}^{K} \frac{\sigma_w\left(I_r^{(n)}\right)}{\sigma_b(n)} N_r + t_b\left(I^{(n)}\right), \\ Q_n^w &:= \sum_{r=1}^{c} \frac{\sigma_b\left(I_r^{(n)}\right)}{\sigma_w(n)} N_r + \sum_{r=c+1}^{K} \frac{\sigma_w\left(I_r^{(n)}\right)}{\sigma_w(n)} N_r + t_w\left(I^{(n)}\right), \end{aligned}$$

with $N_1, \ldots, N_K, I^{(n)}$ independent and $N_1, \ldots, N_K$ standard normally distributed.

The first estimate is via the triangle inequality,

$$\begin{aligned} \zeta_3(Z_n, \mathcal{N}(0,1)) &\leq \zeta_3(Z_n, Q_n^b) + \zeta_3(Q_n^b, \mathcal{N}(0,1)), \\ \zeta_3(\Upsilon_n, \mathcal{N}(0,1)) &\leq \zeta_3(\Upsilon_n, Q_n^w) + \zeta_3(Q_n^w, \mathcal{N}(0,1)). \end{aligned}$$

Here, the summands $\zeta_3(Z_n, Q_n^b)$ and $\zeta_3(\Upsilon_n, Q_n^w)$ give rise to recurrence terms involving $\Delta'(I_r^{(n)})$, similarly as corresponding $\Delta(I_r^{(n)})$ terms appear in (2.8). The crucial part of the present proof is to estimate the summands $\zeta_3(Q_n^b, \mathcal{N}(0,1))$ and $\zeta_3(Q_n^w, \mathcal{N}(0,1))$ sufficiently tight. If we used a direct upper estimate of ζ_3 in terms of ℓ_3, see Lemma 2.1 in [15], and the strategy of proof of the upper bound of Theorem 1.1 in [15], we would only obtain a bound on the rate of convergence of the order $n^{\lambda-\frac{1}{2}}$ for $0 < \lambda < \frac{1}{2}$. To obtain the additional factor 3 in the exponent of our rate $n^{3(\lambda-\frac{1}{2})}$ for $\frac{1}{3} < \lambda < \frac{1}{2}$ and the rate $n^{-\frac{1}{2}+\varepsilon}$ for $0 < \lambda \leq \frac{1}{3}$, we need to bound $\zeta_3(Q_n^b, \mathcal{N}(0,1))$ tighter. We use an idea of [16]: With

$$G_n^b := \left(\sum_{r=1}^{a+1} \frac{\sigma_b^2\left(I_r^{(n)}\right)}{\sigma_b^2(n)} + \sum_{r=a+2}^{K} \frac{\sigma_w^2\left(I_r^{(n)}\right)}{\sigma_b^2(n)} \right)^{\frac{1}{2}},$$

$$A_b := \left\{ G_n^b \geq 1 \right\}, \quad \Delta_n^b := \sqrt{\left| \left(G_n^b\right)^2 - 1 \right|}$$

we have with standard normally distributed N' independent of all other quantities

$$\begin{aligned} Q_n^b &\stackrel{d}{=} \mathbf{1}_{A_b}\left(N + \Delta_n^b N' + t_b\left(I^{(n)}\right)\right) \\ &\quad + \mathbf{1}_{(A_b)^c}\left(G_n^b N + t_b\left(I^{(n)}\right)\right), \\ N &\stackrel{d}{=} \mathbf{1}_{A_b} N + \mathbf{1}_{(A_b)^c}\left(G_n^b + \Delta_n^b N'\right). \end{aligned}$$

Then, a Taylor expansion of the functions $f \in \mathcal{F}_3$ in the definition of ζ_3 and estimates along the ones used on pages 2846–2848 in [16] allow to come up with tighter estimates for $\zeta_3(Q_n^b, \mathcal{N}(0,1))$ (and $\zeta_3(Q_n^w, \mathcal{N}(0,1))$ respectively). The proof of Theorem 1.2 can then be finished by induction.

References

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

[2] Bagchi, A. and Pal, A.K. (1985) Asymptotic Normality in the Generalized Polya–Eggenberger Urn Model, with an Application to Computer Data Structures. *SIAM J. Algebraic Discrete Methods* **6**, 394–405.

[3] Chauvin, B., Mailler, C. and Pouyanne, N. (2015) Smoothing equations for large Pólya urns. *J. Theo. Probab.* **28**, 923–957.

[4] Chow, Y.S. and Teicher, H. (1997) Probability theory. Independence, interchangeability, martingales. Third edition. Springer Texts in Statistics. *Springer-Verlag, New York.*

[5] Fill, J.A. and Janson, S. (2002) Quicksort asymptotics. Analysis of algorithms. *J. Algorithms* **44**, 4–28.

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

[7] Goldstein, L. and Reinert, G. (2013) Stein's method for the Beta distribution and the Pólya-Eggenberger urn. *J. Appl. Probab.* **50**, 1187–1205.

[8] Hwang, H.-K. (2003) Second phase changes in random m-ary search trees and generalized quicksort: convergence rates. *Ann. Probab.* **31**, 609–629.

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

[10] Janson, S. and Pouyanne, N. (2016) Moment convergence of balanced Pólya processes. Preprint, available as arXiv:1606.07022.

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

[12] Knape, M. and Neininger, R. (2014) Pólya urns via the contraction method. *Combin. Probab. Comput.* **23**, 1148–1186.

[13] Kuba, M. and Sulzbach, H. (2016) On martingale tail sums in affine two-color urn models with multiple drawings. *J. Appl. Probab.*, to appear.

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

[15] Neininger, R. and Rüschendorf, L. (2002) Rates of convergence for Quicksort. *J. Algorithms* **44**, 52–62.

[16] Neininger, R. and L. Rüschendorf (2004) On the contraction method with degenerate limit equation. *Ann. Probab.* **32**, 2838–2856.

[17] Peköz, E.A., Röllin, A. and Ross, A. (2016) Generalized Gamma approximation with rates for urns, walks and trees. *Ann. Probab.* **44**, 1776–1816.

[18] Pouyanne, N. (2008) An algebraic approach to Pólya processes. *Ann. Inst. Henri Poincaré Probab. Stat.* **44**, 293–323.

[19] Savkevich, V. (1940) Sur le schéma des urnes a composition variable. *CR (Dokl.) Acad. Sci. URSS* **28**, 8–12.