

RATE OF CONVERGENCE FOR TRADITIONAL PÓLYA URNS

SVANTE JANSON *, Uppsala University

Abstract

Consider a Pólya urn with balls of several colours, where balls are drawn sequentially and each drawn ball is immediately replaced together with a fixed number of balls of the same colour. It is well known that the proportions of balls of the different colours converge in distribution to a Dirichlet distribution. We show that the rate of convergence is $\Theta(1/n)$ in the minimal L_p metric for any $p \in [1, \infty]$, extending a result by Goldstein and Reinert; we further show the same rate for the Lévy distance, while the rate for the Kolmogorov distance depends on the parameters, i.e. on the initial composition of the urn. The method used here differs from the one used by Goldstein and Reinert, and uses direct calculations based on the known exact distributions.

Keywords: Pólya urns; Dirichlet distribution; convergence rate; Wasserstein distance; minimal L_p distance; Kolmogorov–Smirnov distance; Lévy distance

2010 Mathematics Subject Classification: Primary 60C05

Secondary 60F05; 60J99

1. Introduction

A (traditional) Pólya urn contains balls of several colours; we denote the number of colours by q . At discrete time steps, a ball is drawn at random from the urn, and it is replaced together with a balls of the same colour, where $a > 0$ is some given constant. These urn models were studied by Markov in 1917 [11], but they are named after George Pólya who studied them with Eggenberger in the 1920s [2, 15]. (These early references studied the case $q = 2$; the extension to general q is straightforward. See also [7, Chapter 4] and [10], for example.)

Let the vector $\mathbf{X}_n = (X_{n,1}, \dots, X_{n,q})$ ($n \geq 0$) describe the numbers of balls of the different colours at time n . (We assume that the colours are the integers $1, \dots, q$.) The description above thus means that the random vectors \mathbf{X}_n evolve as a Markov chain, with some given (deterministic) initial vector $\mathbf{X}_0 = \mathbf{x}_0 = (x_1, \dots, x_q)$ and

$$\mathbb{P}(\mathbf{X}_{n+1} = \mathbf{X}_n + a\mathbf{e}_i \mid \mathbf{X}_n) = \frac{X_{n,i}}{|\mathbf{X}_n|}, \quad i = 1, \dots, q, \quad (1.1)$$

where $\mathbf{e}_i = (\delta_{ij})_{j=1}^q$ are the unit vectors and $|\mathbf{X}_n| := \sum_{i=1}^q X_{n,i}$. Note that the total number of balls $|\mathbf{X}_n| = an + |\mathbf{x}_0|$ is deterministic.

Although the formulation in the first paragraph above talks about the ‘number of balls’, thus implying that the numbers are integers, we see that no such assumption is needed in the more formal definition using (1.1). Hence, from now on, a and x_1, \dots, x_q may be arbitrary positive real numbers, and thus the random vectors $\mathbf{X}_n \in \mathbb{R}_{>0}^q$. We assume that a and x_1, \dots, x_q are

Received 26 November 2019; revision received 6 May 2020.

* Postal address: Department of Mathematics, Uppsala University, PO Box 480, SE-751 06 Uppsala, Sweden.
svante.janson@math.uu.se

strictly positive to avoid trivial complications; this implies that $X_{n,i} \geq x_i > 0$ for all $n \geq 0$ and $i \in [q] := \{1, \dots, q\}$. In particular, $|\mathbf{X}_n| > 0$, and the transition probabilities are well-defined by (1.1).

One advantage of the extension to real values is that the model is homogeneous in the sense that if we multiply all X_n (including the initial values) and a by the same positive real number, then (1.1) still holds, so we have another Pólya urn. In particular, replacing \mathbf{X}_n with \mathbf{X}_n/a , we may without loss of generality assume that $a = 1$.

We also define the random vector $\mathbf{Y}_n = (Y_{n,1}, \dots, Y_{n,q})$, where $Y_{n,i}$ is the number of the first n drawn balls that have colour i . Obviously

$$\mathbf{X}_n = \mathbf{x}_0 + a\mathbf{Y}_n, \tag{1.2}$$

so it is equivalent to studying \mathbf{X}_n and \mathbf{Y}_n .

It is easy to find the exact distribution of \mathbf{Y}_n and thus of \mathbf{X}_n (see e.g. [11], [12], [15], [7], [10] and (3.3) below), and from this it is easy to see that as $n \rightarrow \infty$, the fraction of balls of a given colour converges in distribution to a beta distribution [15]; more precisely

$$\frac{X_{n,i}}{|\mathbf{X}_n|} \xrightarrow{d} \text{Beta}\left(x_i/a, \sum_{j \neq i} x_j/a\right). \tag{1.3}$$

More generally, the vector $\mathbf{X}_n/|\mathbf{X}_n|$ of proportions converges to a Dirichlet distribution (see Section 2 for the definition):

$$\frac{\mathbf{X}_n}{|\mathbf{X}_n|} \xrightarrow{d} \text{Dir}(\mathbf{x}_0/a). \tag{1.4}$$

It follows by (1.2) that the same holds for the proportions $\mathbf{Y}_n/|\mathbf{Y}_n|$.

Remark 1.1. In fact it is easy to see that $\mathbf{X}_n/|\mathbf{X}_n|$ is a martingale, and thus $\mathbf{X}_n/|\mathbf{X}_n|$ converges a.s. to a limit, which thus has the Dirichlet distribution $\text{Dir}(\mathbf{x}_0/a)$. It follows by (1.2) that the same holds for $\mathbf{Y}_n/|\mathbf{Y}_n|$. The a.s. convergence can also be seen in other ways, for example by de Finetti’s theorem and the fact that the sequence of colours of the drawn balls is exchangeable (see Remark 1.4) or by Martin boundary theory [1, 18].

The purpose of the present paper is to study the rate of convergence of the distributions in (1.3) and (1.4). For the Wasserstein metric d_W , this was done by Goldstein and Reinert [5], who proved, in the case $q = 2$, a bound of the order $O(1/n)$, with an explicit constant depending on a and \mathbf{x}_0 . More recently, Gan, Röllin, and Ross [4] gave, for general q , a bound of the order $O(1/n)$ for a weaker metric. Furthermore, for $q = 2$, $a = 1$, and x_1, x_2 integers, Peköz, Röllin, and Ross, in [13] ($x_1 = 1$) and [14], gave a bound $O(1/n)$ for the stronger metric ℓ_∞ .

The Wasserstein metric equals the minimal L_1 metric ℓ_1 , and our main result is the following, which extends the result by Goldstein and Reinert [5] to ℓ_p for all $p \leq \infty$, and to all numbers of colours $q \geq 2$. See Section 2 for definitions and Sections 3–4 for proofs.

Theorem 1.1. *For any $q \geq 2$, any $a > 0$, and any initial values $\mathbf{X}_0 = \mathbf{x}_0 \in \mathbb{R}_{>0}^q$, if $\mathbf{W} \sim \text{Dir}(\mathbf{x}_0/a)$, then for every $p \in [1, \infty]$ and all $n \geq 1$,*

$$\ell_p\left(\frac{\mathbf{X}_n}{|\mathbf{X}_n|}, \mathbf{W}\right) = \Theta\left(\frac{1}{n}\right), \tag{1.5}$$

$$\ell_p\left(\frac{\mathbf{Y}_n}{n}, \mathbf{W}\right) = \Theta\left(\frac{1}{n}\right). \tag{1.6}$$

Corollary 1.1. For any $q \geq 2$, any $a > 0$, and any initial values $\mathbf{X}_0 = \mathbf{x}_0 \in \mathbb{R}_{>0}^q$, if $W_i \sim \text{Beta}(x_i/a, (|\mathbf{x}_0| - x_i)/a)$, then for every $p \in [1, \infty]$, every $i \in [q]$, and all $n \geq 1$,

$$\ell_p\left(\frac{X_{n,i}}{|\mathbf{X}_n|}, W_i\right) = \Theta\left(\frac{1}{n}\right), \tag{1.7}$$

$$\ell_p\left(\frac{Y_{n,i}}{n}, W_i\right) = \Theta\left(\frac{1}{n}\right). \tag{1.8}$$

We prove Theorem 1.1 by induction on q . The main part of the proof is the base case $q = 2$, which is proved in Section 3. The easy induction for general q is done in Section 4.

The proofs by Goldstein and Reinert [5] and Gan, Röllin, and Ross [4] are based on Stein’s method, where they develop versions for the beta distribution and Dirichlet distribution, respectively. In contrast, the present paper uses only direct calculations, based on the known exact distributions. This is similar in spirit to the proofs by Peköz, Röllin, and Ross [13, 14], which use explicit couplings; however, their method seems difficult to generalize to non-integer x_i/a .

Since the ℓ_p metrics are monotone in p , it suffices to prove the lower bounds for $p = 1$ and the upper bounds for $p = \infty$. Moreover, the lower bounds in (1.5)–(1.8) are trivial, since W_i has a continuous distribution with continuous density function, while $X_{n,i}/|\mathbf{X}_n|$ and $Y_{n,i}/n$ are discrete with values spaced by $a/(an + |\mathbf{x}_0|)$ and $1/n$, respectively; see Section 3 for details. Since $|\mathbf{X}_n| = an + |\mathbf{x}_0|$, the upper bounds in (1.5) and (1.6) are for $p = \infty$ equivalent to, respectively,

$$\ell_\infty(\mathbf{X}_n, |\mathbf{X}_n|\mathbf{W}) = O(1), \tag{1.9}$$

$$\ell_\infty(\mathbf{Y}_n, n\mathbf{W}) = O(1). \tag{1.10}$$

Moreover, by (1.2), (1.9) is equivalent to $\ell_\infty(a\mathbf{Y}_n, an\mathbf{W}) = O(1)$ and thus to (1.10). Hence the upper bounds in the two assertions (1.5) and (1.6) in the theorem are equivalent.

For the one-dimensional variables $X_{n,i}$ and $Y_{n,i}$, we also give the corresponding results for the Kolmogorov–Smirnov distance d_{KS} and the Lévy distance d_{L} . (Upper bounds for a multi-dimensional version of the Kolmogorov–Smirnov distance are given in [4]; the rate obtained there is presumably not sharp, and is weaker than the one found here for the marginals, also when $q = 2$.) The proofs are given in Section 5.

Theorem 1.2. With assumptions and notations as in Corollary 1, and

$$\rho := \min\left\{\frac{x_i}{a}, \frac{|\mathbf{x}_0| - x_i}{a}, 1\right\}, \tag{1.11}$$

we have

$$d_{\text{KS}}\left(\frac{X_{n,i}}{|\mathbf{X}_n|}, W_i\right) = \Theta(n^{-\rho}), \tag{1.12}$$

$$d_{\text{KS}}\left(\frac{Y_{n,i}}{n}, W_i\right) = \Theta(n^{-\rho}). \tag{1.13}$$

Theorem 1.3. With assumptions and notations as in Corollary 1.1,

$$d_{\text{L}}\left(\frac{X_{n,i}}{|\mathbf{X}_n|}, W_i\right) = \Theta(n^{-1}), \tag{1.14}$$

$$d_{\text{L}}\left(\frac{Y_{n,i}}{n}, W_i\right) = \Theta(n^{-1}). \tag{1.15}$$

Remark 1.2. Note that the rate for the Kolmogorov–Smirnov distance differs from the other distances considered here, but only in the case when one of the parameters $\alpha := x_i/a$ and $\beta := (|\mathbf{x}_0| - x_i)/a$ of $W_i \sim \text{Beta}(\alpha, \beta)$ is less than 1. It follows from the proofs below that this difference can be attributed to the fact that the density (2.1) of W_i is unbounded in (precisely) this case.

Remark 1.3. We do not attempt to find explicit values for the constant C in the upper bounds above, although in principle it should be possible by careful calculations in the proof. Note that the proof below treats the cases $x_i < a$, $x_i = a$ and $x_i > a$ separately, and the constants implicit in the calculations might blow up as $x_i \rightarrow a$ from above or below. We conjecture that the constant C can be chosen uniformly for, say, all \mathbf{x}_0 with $a/2 \leq x_i \leq 2a$, but it is not obvious whether this follows from (some version of) our method of proof or not, and we leave that as an open problem.

Remark 1.4. The present paper studies the rate of convergence of the *distribution* of $\mathbf{X}_n/|\mathbf{X}_n|$. As stated above, this sequence of random variables converges a.s. to a limit \mathbf{W} , and one might also ask about the a.s. rate of convergence of the variables. In fact, the sequence of colours of the drawn balls is exchangeable, which (as noted by [15]) follows from the simple formula for the probability distribution of the sequence (see (3.3) and the comments after it). Hence de Finetti’s theorem shows that, conditioned on the a.s. limit \mathbf{W} , the sequence \mathbf{Y}_n is the sequence of partial sums of an independent and identically distributed (i.i.d.) sequence of random unit vectors Z_n , with the distribution $\mathbb{P}(Z_n = \mathbf{e}_i) = W_i$; in particular, each coordinate is the sequence of sums of an i.i.d. Bernoulli sequence. (See e.g. [8, Theorem 11.10].) Consequently we have the same a.s. rate of convergence for the trajectories \mathbf{Y}_n/n and $\mathbf{X}_n/|\mathbf{X}_n|$ as for the strong law of large numbers for i.i.d. Bernoulli sequences. Thus, by the central limit theorem, the distance $|\mathbf{X}_n/|\mathbf{X}_n| - \mathbf{W}|$ is typically of order $n^{-1/2}$, and more precisely there is the law of the iterated logarithm.

Note that the distributions thus converge faster than the trajectories. This method with conditioning on the limit can be used to obtain an upper bound for the distance in (1.6), at least for $p < \infty$ (see [9, Lemma 2.1]), but it only yields a bound $O(n^{-1/2})$, the ℓ_p rate of convergence of $n^{-1} \text{Bi}(n, p')$ to the constant p' , for $p' \in (0, 1)$.

Remark 1.5. Generalized versions of the urn model above, where we may also add to the urn balls of colours other than the colour of the drawn one, have a rich theory with, typically, quite different behaviour. See e.g. [6]. Such urns will not be studied in the present paper.

Remark 1.6. The initial motivation for this study was a paper by Kuntschik and Neininger [9], where generalized Pólya urns (with other replacement schemes) were considered; one step in the proof in [9] was a weaker version of (1.8) with the bound $O(n^{-1/2})$. I initially hoped that the sharper result in Corollary 1.1 would lead to an improvement of the result in [9] for generalized Pólya urns, but it turns out that this is not the case; the estimate found in [9] of the convergence rate is dominated by other terms in their proof.

2. Notation and other preliminaries

2.1. General

For a vector $\mathbf{x} = (x_1, \dots, x_q) \in \mathbb{R}^q$, we define $|\mathbf{x}| := \sum_i |x_i|$. (This is mainly for convenience, in for example (1.1). When defining the metrics below, the Euclidean distance would be just as good, within some constant factors in the result.)

Here $\lfloor x \rfloor$ denotes the integer part of a real number x , and $\{x\} := x - \lfloor x \rfloor$ the fractional part. Here $F_X(x) := \mathbb{P}(X \leq x)$ denotes the distribution function of a real-valued random variable X . Here C and c denote various positive constants that may depend on the parameters q, a and \mathbf{x}_0 , but not on n ; they may change from one occurrence to another.

For sequences a_n and b_n of real numbers, with $b_n > 0$, $a_n = O(b_n)$ means $|a_n| \leq Cb_n$, for some C . Similarly $a_n = \Theta(b_n)$ means $cb_n \leq |a_n| \leq Cb_n$.

2.2. Beta and Dirichlet distributions

Recall that the *beta distribution* $\text{Beta}(\alpha, \beta)$, where $\alpha, \beta > 0$, has the density function

$$f_{\alpha, \beta}(x) = B(\alpha, \beta)^{-1} x^{\alpha-1} (1-x)^{\beta-1} \quad (0 < x < 1), \tag{2.1}$$

where the normalizing constant is given by the *beta function*

$$B(\alpha, \beta) := \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha + \beta)}. \tag{2.2}$$

Recall also that if $q \geq 2$ and $\alpha_1, \dots, \alpha_q > 0$, then the *Dirichlet distribution* $\text{Dir}(\alpha_1, \dots, \alpha_q)$ is the distribution on the simplex $\{(x_1, \dots, x_q) \in \mathbb{R}_{>0}^q : \sum_i x_i = 1\}$ with density

$$\frac{\Gamma(\alpha_1 + \dots + \alpha_q)}{\Gamma(\alpha_1) \dots \Gamma(\alpha_q)} x_1^{\alpha_1} \dots x_q^{\alpha_q} dx_1 \dots dx_{q-1}. \tag{2.3}$$

In the case $q=2$, this is essentially the same as a beta distribution; more precisely $\mathbf{W} \in \text{Dir}(\alpha, \beta)$ if and only if $\mathbf{W} = (V, 1 - V)$, with $V \in \text{Beta}(\alpha, \beta)$. More generally, if $(W_1, \dots, W_q) \in \text{Dir}(\alpha_1, \dots, \alpha_q)$, then $W_i \sim \text{Beta}(\alpha_i, \alpha'_i)$, with $\alpha'_i := \sum_{j \neq i} \alpha_j$.

One well-known construction of Dirichlet-distributed random variables is that if V_1, \dots, V_q are independent random variables with $V_i \sim \Gamma(\alpha_i)$, then the vector of proportions

$$\frac{(V_1, \dots, V_q)}{\sum_i V_i} \sim \text{Dir}(\alpha_1, \dots, \alpha_q), \tag{2.4}$$

and this vector is independent of $\sum_{i=1}^q V_i$. If $q > 2$, then, letting $V' := \sum_{i=2}^q V_i$,

$$\frac{(V_1, \dots, V_q)}{\sum_i V_i} = \left(\frac{V_1}{V_1 + V'}, \frac{V'}{V_1 + V'} \frac{(V_2, \dots, V_q)}{V'} \right), \tag{2.5}$$

and it follows from (2.4) that if $Z \sim \text{Beta}(\alpha_1, \alpha')$ and $\mathbf{V} \sim \text{Dir}(\alpha_2, \dots, \alpha_q)$ are independent, with $\alpha' = \sum_{i=2}^q \alpha_i$, then

$$(Z, (1 - Z)\mathbf{V}) \sim \text{Dir}(\alpha_1, \dots, \alpha_q). \tag{2.6}$$

2.3. Some probability metrics

A probability metric is a metric on a suitable set of probability distributions in some space \mathcal{S} ; in this paper we only consider $\mathcal{S} = \mathbb{R}$ or \mathbb{R}^q for some fixed q . (See e.g. [16] for a general theory and many examples, including the ones below.) Although a probability metric $d(\mu, \nu)$ is formally defined for distributions μ and ν , we follow common practice and write $d(X, Y) := d(\mathcal{L}(X), \mathcal{L}(Y))$ when X and Y are random variables with distributions $\mathcal{L}(X)$ and $\mathcal{L}(Y)$, and we state the definitions below in this form.

In the present paper we mainly study the *minimal L_p distance* ℓ_p for $1 \leq p \leq \infty$, defined by

$$\ell_p(X, Y) := \inf \{ \|X' - Y'\|_p : X' \stackrel{d}{=} X, Y' \stackrel{d}{=} Y \}, \tag{2.7}$$

where $\|X' - Y'\|_p$ is the usual L_p distance, i.e. $(\mathbb{E}|X' - Y'|^p)^{1/p}$ for $p < \infty$, and the essential supremum of $|X' - Y'|$ if $p = \infty$, and the infimum is taken over all *couplings* of X and Y , i.e. all pairs (X', Y') of random variables on a common probability space with the same (marginal) distributions as X and Y . (The infimum is actually attained; see [16, Corollary 5.3.2].) This is a metric on the set of all probability distributions in \mathbb{R}^q with finite p th absolute moment.

The metric ℓ_1 is also known as the *Wasserstein distance* and the *Kantorovich distance*; see [17] for a history.

For random variables X and Y in \mathbb{R} , we also consider the *Kolmogorov–Smirnov distance*

$$d_{KS}(X, Y) := \sup_{x \in \mathbb{R}} |F_X(x) - F_Y(x)| = \sup_{x \in \mathbb{R}} |P(X \leq x) - P(Y \leq x)| \tag{2.8}$$

and the *Lévy distance*

$$d_L(X, Y) := \inf \{ \varepsilon > 0 : F_X(x - \varepsilon) - \varepsilon \leq F_Y(x) \leq F_X(x + \varepsilon) + \varepsilon \text{ for all } x \}. \tag{2.9}$$

3. Proof of Theorem 1.1: $q = 2$

As explained in the Introduction, we may assume without loss of generality that $a = 1$, and we shall do so throughout the proof. For convenience, we call the two colours *white* and *black*, and we write $x_1 = \alpha$ and $x_2 = \beta$. We thus assume $a = 1$ and $\mathbf{x}_0 = (\alpha, \beta)$, and note that then $|\mathbf{X}_n| = n + \alpha + \beta$. Furthermore, $\mathbf{W} = (W_1, W_2) = (W_1, 1 - W_1)$, where W_1 is a random variable with the distribution $\text{Beta}(\alpha, \beta)$,

The parameters α and β are fixed throughout the section. Recall that C and c denote various constants that may depend on α and β , but not on n .

Recalling that $Y_{n,1}$ is the number of white balls drawn in the first n draws, define

$$p_{n,i} := \mathbb{P}(Y_{n,1} = i), \tag{3.1}$$

$$P_{n,k} := \mathbb{P}(Y_{n,1} \leq k) = \sum_{i=0}^k p_{n,i}. \tag{3.2}$$

The basis of the proof is the well-known exact formula [2, 7, 10, 11, 15]

$$\begin{aligned} p_{n,i} &= \binom{n}{i} \frac{\alpha(\alpha + 1) \cdots (\alpha + i - 1)\beta(\beta + 1) \cdots (\beta + n - i - 1)}{(\alpha + \beta)(\alpha + \beta + 1) \cdots (\alpha + \beta + n - 1)} \\ &= \frac{n!}{i!(n - i)!} \frac{\Gamma(\alpha + i)\Gamma(\beta + n - i)\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)\Gamma(\alpha + \beta + n)} \\ &= B(\alpha, \beta)^{-1} \frac{\Gamma(i + \alpha)}{\Gamma(i + 1)} \frac{\Gamma(n - i + \beta)}{\Gamma(n - i + 1)} \frac{\Gamma(n + 1)}{\Gamma(n + \alpha + \beta)} \end{aligned} \tag{3.3}$$

for $i = 0, \dots, n$. This formula is easily verified: if we draw i white balls in the first n draws, then this can be done in $\binom{n}{i}$ ways, and it follows from the definition (1.1) of the urn process that each of them has the same probability given by the fraction on the first line of (3.3). The

rest is simple calculations with gamma functions. For future use we note the special cases

$$p_{n,0} = \frac{\Gamma(\alpha + \beta)}{\Gamma(\beta)} \frac{\Gamma(n + \beta)}{\Gamma(n + \alpha + \beta)}, \tag{3.4}$$

$$p_{n,n} = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)} \frac{\Gamma(n + \alpha)}{\Gamma(n + \alpha + \beta)}. \tag{3.5}$$

Recall that $W_1 \sim \text{Beta}(\alpha, \beta)$ has the density function $f_{\alpha,\beta}$ given in (2.1), and define

$$Q_{n,k} := \mathbb{P}(W_1 \leq k/n) = \int_0^{k/n} f_{\alpha,\beta}(x) dx \tag{3.6}$$

and

$$R_{n,k} := P_{n,k} - Q_{n,k} = \mathbb{P}(Y_{n,1} \leq k) - \mathbb{P}(W_1 \leq k/n). \tag{3.7}$$

We also define $\Delta P_{n,k} := P_{n,k} - P_{n,k-1} = p_{n,k}$, $\Delta Q_{n,k} := Q_{n,k} - Q_{n,k-1}$, and $\Delta R_{n,k} := R_{n,k} - R_{n,k-1}$.

Lemma 3.1.

(i) If $1 \leq k \leq 3n/4$, then

$$|\Delta R_{n,k}| \leq Ck^{\alpha-2}/n^\alpha. \tag{3.8}$$

If further $\alpha = 1$, this can be improved to

$$|\Delta R_{n,k}| \leq Cn^{-2}. \tag{3.9}$$

(ii) If $n/4 \leq k \leq n$, then

$$|\Delta R_{n,k}| \leq C(n - k + 1)^{\beta-2}/n^\beta. \tag{3.10}$$

If further $\beta = 1$, this can be improved to

$$|\Delta R_{n,k}| \leq Cn^{-2}. \tag{3.11}$$

Proof. We recall the standard formula (an easy consequence of Stirling’s formula; see also [12, 5.11.13])

$$\frac{\Gamma(x + a)}{\Gamma(x + b)} = x^{a-b} \left(1 + O\left(\frac{1}{x}\right) \right) \tag{3.12}$$

for, say, fixed $a, b \geq 0$ and $x \geq 1$.

Using (3.12) in (3.3), we find, for $1 \leq k < n$,

$$\begin{aligned} p_{n,k} &= B(\alpha, \beta)^{-1} k^{\alpha-1} (n - k)^{\beta-1} n^{1-\alpha-\beta} \left(1 + O\left(\frac{1}{k}\right) + O\left(\frac{1}{n-k}\right) + O\left(\frac{1}{n}\right) \right) \\ &= n^{-1} f_{\alpha,\beta}\left(\frac{k}{n}\right) \left(1 + O\left(\frac{1}{k}\right) + O\left(\frac{1}{n-k}\right) \right). \end{aligned} \tag{3.13}$$

Furthermore, if $2 \leq k < n$, then, for $x \in ((k - 1)/n, k/n]$,

$$\frac{f_{\alpha,\beta}(x)}{f_{\alpha,\beta}(k/n)} = \left(\frac{nx}{k}\right)^{\alpha-1} \left(\frac{n-nx}{n-k}\right)^{\beta-1} = 1 + O\left(\frac{1}{k}\right) + O\left(\frac{1}{n-k}\right) \tag{3.14}$$

and hence

$$\Delta Q_{n,k} = \int_{(k-1)/n}^{k/n} f_{\alpha,\beta}(x) dx = n^{-1} f_{\alpha,\beta}\left(\frac{k}{n}\right) \left(1 + O\left(\frac{1}{k}\right) + O\left(\frac{1}{n-k}\right)\right). \tag{3.15}$$

Moreover, for $k = 1$, we have, crudely,

$$\Delta Q_{n,1} = \int_0^{1/n} f_{\alpha,\beta}(x) dx \leq C \int_0^{1/n} x^{\alpha-1} dx = Cn^{-\alpha} = n^{-1} f_{\alpha,\beta}\left(\frac{1}{n}\right) O(1), \tag{3.16}$$

and thus (3.15) holds for $k = 1$ as well.

Combining (3.13) and (3.15), we thus obtain, for $1 \leq k < n$,

$$\Delta R_{n,k} = \Delta P_{n,k} - \Delta Q_{n,k} = n^{-1} f_{\alpha,\beta}\left(\frac{k}{n}\right) \left(O\left(\frac{1}{k}\right) + O\left(\frac{1}{n-k}\right)\right). \tag{3.17}$$

If we further assume $1 \leq k \leq 3n/4$, then $f_{\alpha,\beta}(k/n) \leq C(k/n)^{\alpha-1}$, and (3.8) follows. Similar calculations, or interchanging the colours and using (3.8), yield (3.10) for $n/4 \leq k \leq n$.

Now assume $\alpha = 1$. Then, using (3.12), for $k \leq 3n/4$, (3.3) simplifies to

$$p_{n,k} = \beta \frac{\Gamma(n-k+\beta)}{\Gamma(n-k+1)} \frac{\Gamma(n+1)}{\Gamma(n+1+\beta)} = \beta(n-k)^{\beta-1} n^{-\beta} \left(1 + O\left(\frac{1}{n}\right)\right). \tag{3.18}$$

Similarly, still for $k \leq 3n/4$, (3.15) simplifies to

$$\Delta Q_{n,k} = \int_{(k-1)/n}^{k/n} \beta(1-x)^{\beta-1} dx = \beta(n-k)^{\beta-1} n^{-\beta} \left(1 + O\left(\frac{1}{n}\right)\right) \tag{3.19}$$

and we obtain

$$\Delta R_{n,k} = p_{n,k} - \Delta Q_{n,k} = \beta(n-k)^{\beta-1} n^{-\beta} O\left(\frac{1}{n}\right) = O(n^{-2}), \tag{3.20}$$

showing (3.9). The proof of (3.11) is similar, or by interchanging the colours. □

Lemma 3.2.

(i) If $0 \leq k \leq n/2$, then

$$|R_{n,k}| \leq C(k+1)^{\alpha-1} n^{-\alpha}. \tag{3.21}$$

(ii) If $n/2 \leq k \leq n$, then

$$|R_{n,k}| \leq C(n-k+1)^{\beta-1} n^{-\beta}. \tag{3.22}$$

(iii) If $1 \leq k \leq n-1$, then

$$|R_{n,k}| \leq Cn^{-1} f_{\alpha,\beta}(k/n). \tag{3.23}$$

Proof. First consider (i), so $k \leq n/2$. First note that, by (3.4) and (3.12), since $Q_{n,0} = 0$,

$$R_{n,0} = P_{n,0} = p_{n,0} = O(n^{-\alpha}), \tag{3.24}$$

so (3.21) holds for $k = 0$.

We now use Lemma 3.1. If $\alpha > 1$, we have by (3.8), for $1 \leq k \leq n/2$,

$$R_{n,k} = R_{n,0} + \sum_{i=1}^k \Delta R_{n,i} = O\left(n^{-\alpha} + n^{-\alpha} \sum_{i=1}^k i^{\alpha-2}\right) = O(n^{-\alpha} k^{\alpha-1}), \tag{3.25}$$

yielding (3.21) in this case.

If $\alpha = 1$, we obtain (3.21) in the same way, now using (3.9).
 If $\alpha < 1$, (3.8) yields for $0 \leq k \leq \lfloor n/2 \rfloor$ the preliminary result

$$\begin{aligned} R_{n,k} &= R_{n,\lfloor n/2 \rfloor} - \sum_{i=k+1}^{\lfloor n/2 \rfloor} \Delta R_{n,i} \\ &= R_{n,\lfloor n/2 \rfloor} + O\left(n^{-\alpha} \sum_{i=k+1}^{\lfloor n/2 \rfloor} i^{\alpha-2}\right) \\ &= R_{n,\lfloor n/2 \rfloor} + O(n^{-\alpha}(k+1)^{\alpha-1}). \end{aligned} \tag{3.26}$$

It thus remains only to show the case $k = \lfloor n/2 \rfloor$; we postpone this.

Now turn to (ii). If $\beta \geq 1$, we obtain in the same way from (3.10) and (3.11), noting that $R_{n,n} = 1 - 1 = 0$ by the definition (3.7), for $\lfloor n/2 \rfloor \leq k \leq n$,

$$R_{n,k} = - \sum_{i=k+1}^n \Delta R_{n,i} = O(n^{-\beta}(n-k+1)^{\beta-1}). \tag{3.27}$$

If $\beta < 1$, we obtain instead, for $\lfloor n/2 \rfloor \leq k \leq n$,

$$R_{n,k} = R_{n,\lfloor n/2 \rfloor} + \sum_{i=\lfloor n/2 \rfloor+1}^k \Delta R_{n,i} = R_{n,\lfloor n/2 \rfloor} + O(n^{-\beta}(n-k+1)^{\beta-1}). \tag{3.28}$$

We have thus shown, for $0 \leq k \leq \lfloor n/2 \rfloor$,

$$R_{n,k} = \begin{cases} O(n^{-\alpha}(k+1)^{\alpha-1}) & \alpha \geq 1, \\ R_{n,\lfloor n/2 \rfloor} + O(n^{-\alpha}(k+1)^{\alpha-1}) & \alpha < 1, \end{cases} \tag{3.29}$$

and for $\lfloor n/2 \rfloor \leq k \leq n$,

$$R_{n,k} = \begin{cases} O(n^{-\beta}(n-k+1)^{\beta-1}) & \beta \geq 1, \\ R_{n,\lfloor n/2 \rfloor} + O(n^{-\beta}(n-k+1)^{\beta-1}) & \beta < 1. \end{cases} \tag{3.30}$$

This proves (3.21) and (3.22) when $\alpha \geq 1$ and $\beta \geq 1$.

Now suppose that $\alpha < 1$ and $\beta \geq 1$. Then (3.30) yields $R_{n,\lfloor n/2 \rfloor} = O(n^{-1})$, and (3.29) shows that (3.21) holds in this case too.

Similarly, if $\alpha \geq 1$ and $\beta < 1$, then (3.29) yields $R_{n,\lfloor n/2 \rfloor} = O(n^{-1})$ and (3.22) follows from (3.30).

It remains to consider the case $\alpha < 1$ and $\beta < 1$. In this case we sum $R_{n,k}$ over all k and obtain by (3.29) and (3.30), recalling $R_{n,n} = 0$,

$$\begin{aligned} \sum_{k=0}^{n-1} R_{n,k} &= nR_{n,\lfloor n/2 \rfloor} + O\left(n^{-\alpha} \sum_{k=0}^{\lfloor n/2 \rfloor} (k+1)^{\alpha-1}\right) + O\left(n^{-\beta} \sum_{k=\lfloor n/2 \rfloor+1}^{n-1} (n-k)^{\beta-1}\right) \\ &= nR_{n,\lfloor n/2 \rfloor} + O(1). \end{aligned} \tag{3.31}$$

On the other hand, by (3.7),

$$\begin{aligned}
 \sum_{k=0}^{n-1} R_{n,k} &= \sum_{k=0}^{n-1} (\mathbb{P}(Y_{n,1} \leq k) - \mathbb{P}(nW_1 \leq k)) \\
 &= \sum_{k=0}^{n-1} (\mathbb{P}(nW_1 > k) - \mathbb{P}(Y_{n,1} > k)) \\
 &= \sum_{k=0}^{\infty} \mathbb{P}(\lceil nW_1 \rceil > k) - \sum_{k=0}^{\infty} \mathbb{P}(Y_{n,1} > k) \\
 &= \mathbb{E}\lceil nW_1 \rceil - \mathbb{E}Y_{n,1} \\
 &= n\mathbb{E}W_1 + O(1) - \mathbb{E}Y_{n,1}.
 \end{aligned}
 \tag{3.32}$$

However, $\mathbb{E}W_1 = \alpha/(\alpha + \beta)$, and by (1.2), since $\mathbf{X}_n/|\mathbf{X}_n|$ is a martingale and $|\mathbf{X}_n| = n + \alpha + \beta$,

$$\mathbb{E}Y_{n,1} = \mathbb{E}X_{n,1} - \alpha = (n + \alpha + \beta) \frac{\alpha}{\alpha + \beta} - \alpha = n \frac{\alpha}{\alpha + \beta} = n\mathbb{E}W_1.
 \tag{3.33}$$

Hence (3.31)–(3.32) yield $nR_{n,\lfloor n/2 \rfloor} + O(1) = O(1)$, and thus $R_{n,\lfloor n/2 \rfloor} = O(n^{-1})$ also in this case. The proof of (3.21) and (3.22) in this case is now completed by (3.29) and (3.30) as in the cases above.

This completes the proof of (i) and (ii). Finally, (iii) follows from (3.21), (3.22), and (2.1). □

Lemma 3.3. *There exists an integer K such that, for all integers k and $n \geq 1$,*

$$Q_{n,k-K} \leq P_{n,k} \leq Q_{n,k+K}.
 \tag{3.34}$$

Proof. Assume first $n \geq 3L$. For any integer $L \geq 1$, if $L \leq k \leq n - 2L$, then by (2.1),

$$Q_{n,k+L} - Q_{n,k} = \int_{k/n}^{k/n+L/n} f_{\alpha,\beta}(x) dx \geq \frac{L}{n} cf_{\alpha,\beta}\left(\frac{k}{n}\right).
 \tag{3.35}$$

It follows by (3.35) and Lemma 3.2(iii) that if L is chosen large enough, then, for all n and $L \leq k \leq n - 2L$,

$$Q_{n,k+L} - Q_{n,k} \geq R_{n,k} = P_{n,k} - Q_{n,k},
 \tag{3.36}$$

and thus

$$P_{n,k} \leq Q_{n,k+L}.
 \tag{3.37}$$

Furthermore, using (3.37) for $k = L$ (assuming $n \geq 3L$), we see that if $0 \leq k < L$, then

$$P_{n,k} \leq P_{n,L} \leq Q_{n,2L} \leq Q_{n,k+2L}.
 \tag{3.38}$$

Moreover, if $k \geq n - 2L$, then trivially

$$P_{n,k} \leq 1 = Q_{n,n} = Q_{n,k+2L}.
 \tag{3.39}$$

We conclude that if $n \geq 3L$, then for any $k \geq 0$

$$P_{n,k} \leq Q_{n,k+2L}.
 \tag{3.40}$$

Moreover, this is trivially true also if $k < 0$; thus (3.40) holds for all $k \in \mathbb{Z}$ when $n \geq 3L$.

In the same way we see that if L was chosen large enough, also $Q_{n,k} - Q_{n,k-L} \geq -R_{n,k}$ whenever $2L \leq k \leq n - L$, and it follows that $P_{n,k} \geq Q_{n,k-2L}$ for all $k \in \mathbb{Z}$ when $n \geq 3L$.

This proves (3.34) with $K = 3L$ for all $n \geq 1$, since the case $n < 3L$ is trivial. □

Proof of upper bounds in Theorem 1.1 for $q = 2$. We use the monotone coupling of $Y_{n,1}$ and nW_1 . Since nW_1 has a continuous distribution, while $Y_{n,1}$ is discrete, this coupling can be constructed as follows. Define $t_0 < \dots < t_n = n$ such that $\mathbb{P}(nW_1 \leq t_k) = P_{n,k} = \mathbb{P}(Y_{n,1} \leq k)$, and define the function $g: (0, n] \rightarrow \{0, \dots, n\}$ by $g(t) = k$ if $t \in (t_{k-1}, t_k]$ (with $t_{-1} := 0$). Then

$$\mathbb{P}(g(nW_1) \leq k) = \mathbb{P}(nW_1 \leq t_k) = \mathbb{P}(Y_{n,1} \leq k) \quad \text{for } k = 0, \dots, n,$$

and thus $Y_{n,1}^* := g(nW_1)$ has the same distribution as $Y_{n,1}$.

For K as in Lemma 3.3, we have by (3.34) and (3.6)

$$\mathbb{P}(nW_1 \leq k - K) = Q_{n,k-K} \leq P_{n,k} = \mathbb{P}(nW_1 \leq t_k) \leq Q_{n,k+K} = \mathbb{P}(nW_1 \leq k + K), \quad (3.41)$$

and thus $k - K \leq t_k \leq k + K$ for every k . Hence, if $Y_{n,1}^* = k$, so $nW_1 \in (t_{k-1}, t_k]$, then $nW_1 \leq t_k \leq k + K = Y_{n,1}^* + K$, and similarly $nW_1 > t_{k-1} \geq k - 1 - K = Y_{n,1}^* - K - 1$. Consequently, almost surely, $|nW_1 - Y_{n,1}^*| \leq K + 1$, which yields the desired coupling of nW_1 and $Y_{n,1}$. We have thus shown

$$\ell_\infty(Y_{n,1}, nW_1) \leq \|Y_{n,1}^* - nW_1\|_{L^\infty} \leq K_1 := K + 1. \quad (3.42)$$

Since $\mathbf{Y} = (Y_{n,1}, n - Y_{n,1})$ and $\mathbf{W} = (W_1, 1 - W_1)$, this coupling also shows $\ell_\infty(\mathbf{Y}, n\mathbf{W}) \leq K_1$, i.e. (1.10). As stated in the Introduction, by (1.2) this is equivalent to (1.9), and the upper bounds in (1.5)–(1.6) follow. □

For the proof of the lower bounds, we record a simple general fact.

Lemma 3.4. *If W is a continuous random variable, then $\{nW\} \xrightarrow{d} U(0, 1)$ as*

$n \rightarrow \infty$. In particular,

$$\mathbb{P}\left(\frac{1}{4} < \{nW\} < \frac{3}{4}\right) \rightarrow \frac{1}{2} \quad \text{as } n \rightarrow \infty. \quad (3.43)$$

Proof. As is well known, by considering the Fourier transform on $\mathbb{T} = \mathbb{R}/\mathbb{Z}$, $\{nW\} \xrightarrow{d} U(0, 1)$ is equivalent to

$$\mathbb{E} e^{2\pi im\{nW\}} \rightarrow 0 \quad \text{as } n \rightarrow \infty \quad (3.44)$$

for every $m \in \mathbb{Z} \setminus \{0\}$. We have

$$\mathbb{E} e^{2\pi im\{nW\}} = \mathbb{E} e^{2\pi imnW} = \varphi(2\pi mn), \quad (3.45)$$

where φ is the characteristic function of W , and thus (3.44) follows by the Riemann–Lebesgue lemma. □

Proof of lower bounds in Theorem 1.1 and Corollary 1.1. (This proof holds for any $q \geq 2$.) Note that W_1 is a continuous random variable while $Y_{n,1}$ is integer-valued. Hence, in any coupling of $Y_{n,1}$ and W_1 ,

$$\mathbb{P}\left(|Y_{n,1} - nW_1| > \frac{1}{4}\right) \geq \mathbb{P}\left(\frac{1}{4} < \{nW_1\} < \frac{3}{4}\right) \geq c \quad (3.46)$$

for some $c > 0$ and all $n \geq 1$, since the last probability in (3.46) is strictly positive for every n and converges to $\frac{1}{2}$ as $n \rightarrow \infty$ by Lemma 3.4. It follows that $\ell_p(Y_{n,1}, nW_1) \geq \ell_1(Y_{n,1}, nW_1) \geq c/4$. The lower bound in (1.8) follows. The proof of the lower bound in (1.7) is similar, and the lower bounds in (1.5) and (1.6) follow trivially from these. \square

4. Proof of Theorem 1.1: $q > 2$

We show here how the general case of Theorem 1.1 follows from the case $q = 2$ shown in Section 3. Recall that the lower bounds were proved for general q in (3.46). Hence it suffices to show the upper bounds.

Proof of upper bounds in Theorem 1.1 for $q > 2$. We may still assume $a = 1$ for simplicity. We focus on the version (1.10); this implies (1.9) and (1.5)–(1.6), as stated in the Introduction.

We use induction on q . Thus, let $q > 2$, and assume that (1.10) holds when the number of colours is $q - 1$, and when it is 2.

We call the first colour *white* and all other colours *dark*. It is then obvious from the definition of the urn, in words or by (1.1), that if we are colour-blind and only notice whether balls are white or dark, we obtain another Pólya urn, with two colours. Formally, we define $\mathbf{X}'_n := (X_{n,2}, \dots, X_{n,q})$ and $X'_n := |\mathbf{X}'_n|$, and then $(X_{n,1}, X'_n)$ is a Pólya urn with initial state $\mathbf{x}'_0 = (x_1, x')$ where $x' := \sum_{j \neq 1} x_j$, and the same $a = 1$.

Let also $\mathbf{Y}'_n := (Y_{n,2}, \dots, Y_{n,q}) = \mathbf{X}'_n - \mathbf{x}'_0$ and $Y'_n := |\mathbf{Y}'_n|$, the number of times in the n first draws that a dark ball is drawn. Then it follows from the case $q = 2$ of (1.10) applied to the urn with white and dark balls that if $Z \sim \text{Beta}(x_1, x')$, then, for each $n \geq 1$,

$$\ell_\infty(Y_{n,1}, nZ) \leq C. \tag{4.1}$$

In other words, there exists a random variable $Z^* \sim \text{Beta}(x_1, x')$ such that a.s. $|Y_{n,1} - nZ^*| \leq C$, that is,

$$Y_{n,1} = nZ^* + O_{L_\infty}(1), \tag{4.2}$$

where, as in the rest of this proof, $O_{L_\infty}(1)$ denotes a random variable that is a.s. bounded by some constant C . (These random variables, and other variables in the proof such as Z^* , may depend on n , but the bound C does not.) Note also that $Y_{n,1} + Y'_n = |\mathbf{Y}_n| = n$, and thus

$$Y'_n = n - Y_{n,1} = n(1 - Z^*) + O_{L_\infty}(1). \tag{4.3}$$

Moreover, it also follows from the definition that, conditioned on $Y'_n = m$, the vector \mathbf{X}'_n is distributed as the m th stage of a Pólya urn with $q - 1$ colours and initial state \mathbf{X}'_0 . Consequently, using the induction hypothesis on this urn, if $\mathbf{W}' \sim \text{Dir}(x_2, \dots, x_q)$, then

$$\ell_\infty((\mathbf{Y}'_n | Y'_n = m), m\mathbf{W}') \leq C.$$

This means that for every $m \in \{0, \dots, n\}$ there exists a random vector \mathbf{W}_m^* defined on the event $\{Y'_n = m\}$ such that

$$(\mathbf{W}_m^* | Y'_n = m) \stackrel{d}{=} \mathbf{W}' \tag{4.4}$$

and, if $Y'_n = m$, then a.s.

$$|\mathbf{Y}'_n - m\mathbf{W}_m^*| \leq C. \tag{4.5}$$

Define the random vector \mathbf{W}^* on our probability space Ω by $\mathbf{W}^* := \mathbf{W}_{Y'_n}^*$, i.e. $\mathbf{W}^* = \mathbf{W}_m^*$ when $Y'_n = m$. Then (4.4) means that

$$\mathbf{W}^* \stackrel{d}{=} \mathbf{W}' \sim \text{Dir}(x_2, \dots, x_q),$$

and that \mathbf{W}^* is independent of Y'_n , while (4.5) means that $|\mathbf{Y}'_n - Y'_n \mathbf{W}^*| \leq C$, or equivalently

$$\mathbf{Y}'_n = Y'_n \mathbf{W}^* + O_{L_\infty}(1). \tag{4.6}$$

Note also that since \mathbf{W}^* is independent of Y'_n , and thus of $Y_{n,1}$, we may assume that Z^* in (4.2)–(4.3) is chosen to be independent of \mathbf{W}^* .

Combining (4.6) and (4.2)–(4.3), we obtain, recalling that the Dirichlet distribution is supported on a bounded set,

$$\begin{aligned} \mathbf{Y}_n &= (Y_{1,n}, \mathbf{Y}'_n) = (Y_{1,n}, Y'_n \mathbf{W}^*) + O_{L_\infty}(1) \\ &= (nZ^*, n(1 - Z^*)\mathbf{W}^*) + O_{L_\infty}(1) \\ &= n(Z^*, (1 - Z^*)\mathbf{W}^*) + O_{L_\infty}(1). \end{aligned} \tag{4.7}$$

Since $Z^* \sim \text{Beta}(x_1, x')$ and $\mathbf{W}^* \sim \text{Dir}(x_2, \dots, x_q)$, with Z^* and \mathbf{W}^* independent,

$$\mathbf{W} := (Z^*, (1 - Z^*)\mathbf{W}^*) \sim \text{Dir}(x_1, \dots, x_q)$$

by (2.6). Consequently (4.7) shows that $\ell_\infty(\mathbf{Y}_n, n\mathbf{W}) \leq \|\mathbf{Y}_n - n\mathbf{W}\|_{L_\infty} = O(1)$ with $\mathbf{W} \sim \text{Dir}(x_1, \dots, x_q)$, which completes the induction.

This completes the proof of Theorem 1.1. □

Proof of Corollary 1.1. The upper bounds follow by Theorem 1.1, and the lower bounds were proved above by (3.46) and the corresponding argument for $X_{n,i}$. □

5. Proofs of Theorems 1.2 and 1.3

Proof of Theorem 1.2. We may assume $a = 1$, and by symmetry, it suffices to consider $i = 1$. Furthermore, by regarding the colours as white or dark as in Section 4, it suffices to consider $q = 2$. As in Section 3, let $\alpha := x_1$ and $\beta := |\mathbf{x}_0| - x_1$, so (1.11) becomes

$$\rho = \min\{\alpha, \beta, 1\}. \tag{5.1}$$

For simplicity, we consider only (1.13). Similar arguments yield (1.12); note that by (1.2),

$$\begin{aligned} d_{\text{KS}}\left(\frac{X_{n,1}}{|\mathbf{X}_n|}, W_1\right) &= d_{\text{KS}}\left(\frac{Y_{n,1} + x_1}{n + |\mathbf{x}_0|}, W_1\right) \\ &= d_{\text{KS}}\left(\frac{Y_{n,1} + x_1}{n}, \frac{n + |\mathbf{x}_0|}{n} W_1\right) \\ &= d_{\text{KS}}\left(\frac{Y_{n,1}}{n}, \left(1 + \frac{|\mathbf{x}_0|}{n}\right) W_1 - \frac{x_1}{n}\right). \end{aligned} \tag{5.2}$$

Since $Y_{n,1}$ is integer-valued, we have, for $x \in [k/n, (k + 1)/n]$ with integer k , recalling the definitions (3.2), (3.6), (3.7),

$$\begin{aligned} \mathbb{P}(Y_{n,1}/n \leq x) - \mathbb{P}(W_1 \leq x) &= \mathbb{P}(Y_{n,1} \leq k) - \mathbb{P}(W_1 \leq x) \\ &= R_{n,k} - \int_{k/n}^x f_{\alpha,\beta}(x) dx. \end{aligned} \tag{5.3}$$

The integral in (5.3) is less than $\Delta Q_{n,k+1}$, and thus, recalling also (2.8),

$$d_{KS}(Y_{n,1}/n, W_1) \leq \sup_k |R_{n,k}| + \sup_k |\Delta Q_{n,k}|. \tag{5.4}$$

Furthermore, by taking $x = k/n$ in (5.3),

$$\sup_k |R_{n,k}| \leq d_{KS}(Y_{n,1}/n, W_1), \tag{5.5}$$

and by instead letting $x \nearrow (k + 1)/n$,

$$\sup_k |R_{n,k} - \Delta Q_{n,k+1}| \leq d_{KS}(Y_{n,1}/n, W_1). \tag{5.6}$$

Hence, combining (5.5) and (5.6),

$$\sup_k |\Delta Q_{n,k+1}| \leq 2d_{KS}(Y_{n,1}/n, W_1). \tag{5.7}$$

It follows from (3.6) and (2.1) that

$$\begin{aligned} \Delta Q_{n,1} &\geq cn^{-\alpha}, \\ \Delta Q_{n,n} &\geq cn^{-\beta}, \\ \Delta Q_{n, \lfloor (n+1)/2 \rfloor} &\geq c/n. \end{aligned}$$

Consequently (5.7) yields the lower bound, using (5.1),

$$d_{KS}(Y_{n,1}/n, W_1) \geq c \max(n^{-\alpha}, n^{-\beta}, n^{-1}) = cn^{-\rho}. \tag{5.8}$$

For an upper bound, we similarly have (see (3.15)–(3.16))

$$\max_k \Delta Q_{n,k} \leq C \max_{1 \leq k \leq n-1} n^{-1} f_{\alpha,\beta} \left(\frac{k}{n} \right) \leq Cn^{-\rho}. \tag{5.9}$$

Furthermore, Lemma 3.2 yields

$$\max_k |R_{n,k}| \leq Cn^{-\min\{\alpha,\beta,1\}} = Cn^{-\rho}. \tag{5.10}$$

Consequently the upper bound in (1.13) follows from (5.4). □

Proof of Theorem 1.3. We use the same simplifying assumptions and notations as in the proof of Theorem 1.2. Again we consider only $Y_{n,1}/n$; the proof for $X_{n,1}/|\mathbf{X}_n|$ is similar. Lemma 3.3 implies that

$$d_L \left(\frac{Y_{n,1}}{n}, W_1 \right) \leq \frac{K}{n}. \tag{5.11}$$

Conversely, suppose that $d_L(Y_{n,1}/n, W_1) < 1/4n$. Take any δ such that $d_L(Y_{n,1}/n, W_1) < \delta < 1/4n$, and let $w_0 := \lfloor n/2 \rfloor/n$. Then, by (2.9) and the fact that $Y_{n,1}$ takes only integer values,

$$\begin{aligned} F_{W_1} \left(w_0 + \frac{3}{4n} \right) &\leq F_{Y_{n,1}/n} \left(w_0 + \frac{3}{4n} + \delta \right) + \delta \\ &= F_{Y_{n,1}/n} \left(w_0 + \frac{1}{4n} - \delta \right) + \delta \\ &\leq F_{W_1} \left(w_0 + \frac{1}{4n} \right) + 2\delta. \end{aligned} \tag{5.12}$$

We have $W_1 \sim \text{Beta}(\alpha, \beta)$ for some $\alpha, \beta > 0$, and thus (2.1) yields

$$F_{W_1}\left(w_0 + \frac{3}{4n}\right) - F_{W_1}\left(w_0 + \frac{1}{4n}\right) = \int_{w_0+1/(4n)}^{w_0+3/(4n)} f_{\alpha,\beta}(x) dx \geq \frac{c}{n}. \quad (5.13)$$

Together, (5.12) and (5.13) yield $2\delta \geq c/n$. We may choose δ arbitrarily close to $d_L(Y_{n,1}/n, W_1)$, and thus it follows that if

$$d_L(Y_{n,1}/n, W_1) < \frac{1}{4n},$$

then

$$d_L(Y_{n,1}/n, W_1) \geq \frac{c}{2n}.$$

Consequently, in any case,

$$d_L\left(\frac{Y_{n,1}}{n}, W_1\right) \geq \frac{\min\{1/4, c/2\}}{n}. \quad (5.14)$$

This together with (5.11) completes the proof. \square

Acknowledgement

This work was partly carried out during a visit to the Isaac Newton Institute for Mathematical Sciences during the programme ‘Theoretical Foundations for Statistical Network Analysis’ in 2016 (EPSCR grant EP/K032208/1) and was partially supported by a grant from the Simons foundation and a grant from the Knut and Alice Wallenberg Foundation. I thank Ralph Neininger for posing the problem, and Gesine Reinert and two anonymous referees for helpful comments.

References

- [1] BLACKWELL, D. AND KENDALL, D. (1964). The Martin boundary of Pólya’s urn scheme, and an application to stochastic population growth. *J. Appl. Prob.* **1**, 284–296.
- [2] EGGENBERGER, F. AND G. PÓLYA, G. (1923). Über die Statistik verketteter Vorgänge. *Z. Angew. Math. Mech.* **3**, 279–289.
- [3] FREEDMAN, D. A. (1965). Bernard Friedman’s urn. *Ann. Math. Statist.* **36**, 956–970.
- [4] GAN, H. L., RÖLLIN, A. AND ROSS, N. (2017). Dirichlet approximation of equilibrium distributions in Cannings models with mutation. *Adv. Appl. Prob.* **49**, 927–959.
- [5] GOLDSTEIN, L. AND REINERT, G. (2013). Stein’s method for the beta distribution and the Pólya–Eggenberger urn. *J. Appl. Prob.* **50**, 1187–1205.
- [6] JANSON, S. (2004). Functional limit theorems for multitype branching processes and generalized Pólya urns. *Stoch. Process. Appl.* **110**, 177–245.
- [7] JOHNSON, N. L. AND KOTZ, S. (1977). *Urn Models and Their Application*. John Wiley, New York.
- [8] KALLENBERG, O. (2002). *Foundations of Modern Probability*, 2nd edn. Springer, New York.
- [9] KUNTSHIK, A. AND NEININGER, R. (2017). Rates of convergence for balanced irreducible two-color Pólya urns. In *2017 Proceedings of the Fourteenth Workshop on Analytic Algorithmics and Combinatorics (ANALCO)*, pp. 94–99. SIAM, Philadelphia, PA.
- [10] MAHMOUD, H. M. (2009). *Pólya Urn Models*. CRC Press, Boca Raton, FL.
- [11] MARKOV, A. A. (1917). Sur quelques formules limites du calcul des probabilités (Russian). *Bulletin de l’Académie Impériale des Sciences* **11**, 177–186.
- [12] OLVER, F. W. J., LOZIER, D. W., BOISVERT, R. F. AND CLARK, C. W. (eds) (2010). *NIST Handbook of Mathematical Functions*, Cambridge University Press. Also available as *NIST Digital Library of Mathematical Functions*, <http://dlmf.nist.gov/>.
- [13] PEKÖZ, E., RÖLLIN, A. AND ROSS, N. (2016). Generalized gamma approximation with rates for urns, walks and trees. *Ann. Prob.* **44**, 1776–1816.

- [14] PEKÖZ, E., RÖLLIN, A. AND ROSS, N. (2017). Joint degree distributions of preferential attachment random graphs. *Adv. Appl. Prob.* **49**, 368–387.
- [15] PÓLYA, G. (1930). Sur quelques points de la théorie des probabilités. *Ann. Inst. H. Poincaré* **1**, 117–161.
- [16] RACHEV, S. T., KLEBANOV, L. B., STOYANOV, S. V. AND FABOZZI, F. J. (2013). *The Methods of Distances in the Theory of Probability and Statistics*. Springer, New York.
- [17] RÜSCHENDORF, L. Wasserstein metric. *Encyclopedia of Mathematics*. Available at https://www.encyclopediaofmath.org/index.php?title=Wasserstein_metric.
- [18] SAWYER, S. A. (1997). Martin boundaries and random walks. *Harmonic Functions on Trees and Buildings (New York, 1995)* (Contemp. Math. **206**), pp. 17–44. American Mathematical Society, Providence, RI.