

STEIN'S METHOD FOR THE BETA DISTRIBUTION AND THE PÓLYA–EGGENBERGER URN

LARRY GOLDSTEIN,* *University of Southern California*

GESINE REINERT,** *University of Oxford*

Abstract

Using a characterizing equation for the beta distribution, Stein's method is applied to obtain bounds of the optimal order for the Wasserstein distance between the distribution of the scaled number of white balls drawn from a Pólya–Eggenberger urn and its limiting beta distribution. The bound is computed by making a direct comparison between characterizing operators of the target and the beta distribution, the former derived by extending Stein's density approach to discrete distributions. In addition, refinements are given to Döbler's (2012) result for the arcsine approximation for the fraction of time a simple random walk of even length spends positive, and so also to the distributions of its last return time to 0 and its first visit to its terminal point, by supplying explicit constants to the present Wasserstein bound and also demonstrating that its rate is of the optimal order.

Keywords: Stein's method; urn model; beta distribution; arcsine distribution

2010 Mathematics Subject Classification: Primary 60F05

Secondary 62E20; 60K99

1. Introduction

The classical Pólya–Eggenberger urn at time 0 contains $\alpha \geq 1$ white balls and $\beta \geq 1$ black balls, and at every positive integer time a ball is chosen uniformly from the urn, independently of the past, and replaced along with $m \geq 1$ additional balls of the same color. With \mathcal{L} indicating distribution, or law, and \xrightarrow{D} indicating convergence in distribution, it is well known (see, e.g. [23]) that if $S_n = S_n^{\alpha, \beta, m}$ is the number of white balls drawn from the urn by time $n = 0, 1, 2, \dots$ then, as $n \rightarrow \infty$,

$$\mathcal{L}(W_n) \xrightarrow{D} \mathcal{B}\left(\frac{\alpha}{m}, \frac{\beta}{m}\right), \quad \text{where } W_n = \frac{S_n}{n}. \tag{1.1}$$

Here, for positive real numbers α and β , we let $\mathcal{B}(\alpha, \beta)$ denote the beta distribution having density

$$p(x; \alpha, \beta) = \frac{x^{\alpha-1}(1-x)^{\beta-1}}{B(\alpha, \beta)} \mathbf{1}_{\{x \in [0, 1]\}}, \tag{1.2}$$

where $B(\alpha, \beta) = \Gamma(\alpha)\Gamma(\beta)/\Gamma(\alpha + \beta)$ is the beta function as expressed in terms of the gamma function $\Gamma(x)$.

Using Stein's method, we derive an order $O(1/n)$ bound in the Wasserstein distance d_W , defined in (3.1), between W_n and its limiting beta distribution in (1.1). We show in Remark 3.1

Received 13 July 2012; revision received 2 January 2013.

* Postal address: Department of Mathematics, University of Southern California, KAP 108, Los Angeles, CA 90089-2532, USA. Email address: larry@math.usc.edu

** Postal address: Department of Statistics, University of Oxford, 1 South Parks Road, Oxford OX1 3TG, UK. Email address: reinert@stats.ox.ac.uk

that the rate of Theorem 1.1 cannot be improved. We let $x \vee y$ or $\max(x; y)$ denote the maximum of two real numbers x and y , and $x \wedge y$ or $\min(x; y)$ their minimum.

Theorem 1.1. *For $\alpha \geq 1$ and $\beta \geq 1$, let S_n be the number of white balls in n draws from a Pólya–Eggenberger urn that initially contains α white balls and β black balls. Then, with $W_n = S_n/n$ and $Z \sim \mathcal{B}(\alpha/m, \beta/m)$,*

$$d_W(W_n, Z) \leq \left(\frac{m + \alpha \vee \beta}{2nm} + \frac{\alpha\beta}{nm(\alpha + \beta)} \right) (b_0 + b_1) + \frac{3}{2n},$$

where $b_0 = b_0(\alpha/m, \beta/m)$ and $b_1 = b_1(\alpha/m, \beta/m)$ are given in Lemma 3.4.

Connections between Theorem 1.1 and Döbler’s work [12] are spelled out in Remark 3.2.

The $\mathcal{B}(\frac{1}{2}, \frac{1}{2})$ distribution, also known as the arcsine law, describes the asymptotic distribution of many quantities that arise naturally in the study of the simple symmetric random walk $T_n = X_1 + \dots + X_n$, where X_1, \dots, X_n are independent variables taking the values 1 and -1 with probability $\frac{1}{2}$. For instance, let L_{2n} be the random variable

$$L_{2n} = \sup\{m \leq 2n : T_m = 0\}$$

giving the last return time to 0 up to time $2n$. Then, see [15],

$$\mathbb{P}(L_{2n} = 2k) = u_{2k}u_{2n-2k}, \quad \text{where } u_{2m} = \mathbb{P}(T_{2m} = 0) \tag{1.3}$$

with $\mathbb{P}(T_{2m} = 0) = 2^{-2m} \binom{2m}{m}$, the probability that the walk returns to 0 at time $2m$.

In the limit, $(2n)^{-1}L_{2n} \rightarrow Z$ in probability, where Z has the arcsine distribution. It is often noted that this limiting result is somewhat counterintuitive in that the arcsine density has greatest mass near the endpoints, and least mass in the center of the unit interval, whereas in a fair coin tossing game one might assume that players are more likely to spend equal time in the lead. Perhaps at least as remarkable is the fact that the number U_{2n} of segments of the walk that lie above the x -axis, and R_{2n} , the first time the walk visits the terminal point S_{2n} , are all equal in distribution to $\mathcal{L}(L_{2n})$; see [15] for a nice exposition.

Using the methods presented here that were available in a preprint of this article, Döbler [11] presented a Wasserstein bound of order $O(1/n)$ without explicit constants between the distribution of $(2n)^{-1}U_{2n}$ and the limiting arcsine. In Section 4, essentially by applying the bounds in Lemma 3.4, we are able to attach concrete constants to the result of [11], as well as show the rate of the bound is optimal.

Theorem 1.2. *Let L_{2n} be the last return time to 0 of a simple symmetric random walk of length $2n$, and let Z have the arcsine distribution. Then*

$$d_W\left(\frac{L_{2n}}{2n}, Z\right) \leq \frac{27}{2n} + \frac{8}{n^2}.$$

The same bound holds with L_{2n} replaced by U_{2n} or R_{2n} . The $O(1/n)$ rate of the bound cannot be improved.

Beginning with the introduction by Stein [33] of a ‘characterizing equation’ method for developing bounds in normal approximation, to date the method has been successfully applied to a large number of the classical distributions, including the Poisson [2], [4], multinomial [20], gamma [21], [24], geometric [26], negative binomial [5], and exponential [6], [27], [28], as well

as the classical generalized gamma distributions [30] and the nonclassical PRR family [29], both based on Pólya-type urn models. Here we further extend the range of Stein's method by including the beta distribution, focusing on its role as the limiting law of the fraction of white balls drawn from the Pólya–Eggenberger urn.

The application of Stein's method here differs from the way it is usually applied in that we focus on the approximation of particular distributions whose exact forms are known, rather than develop a bound that applies to a class of complex distributions obtained by, say, summing random variables that obey weak moment and dependence conditions, as in the case of the central limit theorem. And indeed, though explicit formulae exist for the distributions we study, the need for their approximation arises regardless, as is the case also for, say, the ubiquitous use of the normal approximation for the binomial.

Urn models of the classical type, and generalizations including drawing multiple balls or starting new urns, have received considerable attention recently; see, for example, [1], [7], and [10]. Interest has partly been sparked by the ability of urn models to exhibit power-law limiting behavior, which in turn has been a focus of network analysis; see, for example, [13] and [29]. Connections between urn models and binary search trees are clearly explained in [22]. In particular, let $m = 1$ and consider the initial state of the Pólya–Eggenberger urn as a rooted binary tree having α white leaves and β black leaves, or external nodes. At every time step one external node is chosen, uniformly, to duplicate, yielding a pair of leaves of the same color. That is, the chosen external node becomes an internal node while two external nodes of the chosen color are added. The rule for adding an additional white leaf to the tree at time n clearly is the same as the rule for adding an additional white ball to the Pólya–Eggenberger urn for the case $m = 1$, and, hence, the number of white leaves of the tree and white balls in the urn have the same distribution. Many variations and extensions on this theme are possible. Another line of interest comes from edge reinforced random walks, because an infinite system of independent Pólya–Eggenberger urns can be used to represent edge reinforced random walks on trees; see [31].

2. Characterizing equations and generators

Stein's method for distributional approximation is based on a characterization of the target approximating distribution. For the seminal normal case considered in [33], it was shown that a variable Z has the standard normal distribution if and only if

$$\mathbb{E}[Zf(Z)] = \mathbb{E}[f'(Z)] \quad (2.1)$$

for all absolutely continuous functions f for which $\mathbb{E}|f'(Z)| < \infty$. If a variable W has an approximate normal distribution, then one expects that it satisfies (2.1) approximately. More specifically, if we wish to test the difference between the distribution of W and the standard normal Z on a function h , then instead of computing $\mathbb{E}h(W) - Nh$, where $Nh = \mathbb{E}h(Z)$, we can set up a 'Stein equation'

$$f'(w) - wf(w) = h(w) - Nh \quad (2.2)$$

for the given h , solve for $f(w)$, and, upon replacing w by W in (2.2), calculate the expectation of the right-hand side by taking the expectation on the left. At first glance it may seem that doing so does not make the given problem any less difficult. However, a number of techniques may be brought to bear on the quantity $\mathbb{E}[f'(W) - Wf(W)]$. In particular, this expression contains only the single random variable W , in contrast to the difference of the expectations of $h(W)$ and $h(Z)$, depending on two distributions.

To obtain our result, we compute the distance between the distribution of the fraction of white balls drawn from the Pólya–Eggenberger urn and the beta by comparing the operators that characterize them. Our approach in characterizing the urn distribution stems from what is known as the density method; see, for instance, [8, Section 13.1], [32], or [34]. In particular, recognizing the $-w$ in (2.2) as the ratio $\phi'(w)/\phi(w)$, where $\phi(w)$ is the standard normal density, one hopes to replace the term $-w$ by the ratio $p'(w)/p(w)$ when developing the Stein equation to handle the distribution with density $p(w)$, and to apply similar reasoning when the distribution under study is discrete. Use of the density method in the discrete case, followed by the application of a judiciously chosen transformation, leads to the characterization of the Pólya–Eggenberger urn distribution given in Lemma 2.1.

Another approach to construct characterizing equations is known as the generator method. A number of years following the publication of [33], the relationship between the characterizing equation (2.1) and the generator of the Ornstein–Uhlenbeck process

$$\mathcal{A}f(w) = f''(w) - wf'(w),$$

of which the normal is the unique stationary measure, was recognized in [3], where it was noted that in some generality the process semigroup may be used to solve the Stein equation (2.2). Given this connection between Stein characterizations and generators, it is natural to consider a stochastic process which has the given target as its stationary distribution when extending Stein’s method to handle a new distribution.

Regarding the use of this ‘generator’ method for extending the scope of Stein’s method to the beta distribution, we recall that the Fisher–Wright model from genetics, originating in [16], [36], and [37], is a stochastic process used to model genetic drift in a population and has a generator given by

$$\mathcal{A}f(x) = w(1 - w)f''(w) + (\alpha(1 - w) - \beta w)f'(w)$$

for positive α and β , and that the $\mathcal{B}(\alpha, \beta)$ distribution is its unique stationary distribution. In particular, with $Z \sim \mathcal{B}(\alpha, \beta)$ we have $\mathbb{E}\mathcal{A}f(Z) = 0$. Let $\mathcal{B}_{\alpha, \beta}h = \mathbb{E}h(Z)$, the $\mathcal{B}(\alpha, \beta)$ expectation of a function h ; we drop the subscripts when the role of the parameters α and β is clear. As $\mathbb{E}h(Z) - \mathcal{B}h$ is also 0, we are led to consider a Stein equation for the beta distribution of the form

$$w(1 - w)f'(w) + (\alpha(1 - w) - \beta w)f(w) = h(w) - \mathcal{B}_{\alpha, \beta}h. \tag{2.3}$$

Lemma 2.1 provides a characterizing equation for the Pólya urn distribution that is parallel to (2.3). Taking differences then allows us to estimate the expectation of the right-hand side of (2.3) when w is replaced by W_n by exploiting the similarity of the two characterizing operators; a similar argument can be found in [14] and [18] for stationary distributions of birth–death chains. The results most closely related to the present work are those of [12], and their connections to the present manuscript are discussed in Remark 3.2.

First we introduce some notation. We say a subset I of the integers \mathbb{Z} is a finite integer interval if $I = [a, b] \cap \mathbb{Z}$ for $a, b \in \mathbb{Z}$ with $a \leq b$, and an infinite integer interval if either $I = (-\infty, b] \cap \mathbb{Z}$ or $I = [a, \infty) \cap \mathbb{Z}$ or $I = \mathbb{Z}$. For a real-valued function f , let $\Delta f(k) = f(k + 1) - f(k)$, the forward difference operator, and, for a real-valued function p taking nonzero values in the integer interval I , let

$$\psi(k) = \frac{\Delta p(k)}{p(k)} \quad \text{for } k \in I. \tag{2.4}$$

For Z a random variable having probability mass function p with support the integer interval I , let $\mathcal{F}(p)$ denote the set of all real-valued functions f such that either $\mathbb{E}\Delta f(Z - 1)$ or $\mathbb{E}\psi(Z)f(Z)$ is finite, $\lim_{n \rightarrow \infty} f(n)p(n + 1) = 0$ when $\sup\{k : p(k) > 0\} = \infty$, $\lim_{n \rightarrow -\infty} f(n)p(n + 1) = 0$ when $\inf\{k : p(k) > 0\} = -\infty$, and $f(a - 1) = 0$ in the case where $I = [a, b] \cap \mathbb{Z}$ or $I = [a, \infty) \cap \mathbb{Z}$.

Lemma 2.1. *Let p be the probability mass function of the number $S_n^{\alpha, \beta, m}$ of white balls drawn from the Pólya–Eggenberger urn by time n . Then a random variable S has probability mass function p if and only if*

$$\mathbb{E}\left[S\left(\frac{\beta}{m} + n - S\right)\Delta f(S - 1) + \left\{(n - S)\left(\frac{\alpha}{m} + S\right) - S\left(\frac{\beta}{m} + n - S\right)\right\}f(S)\right] = 0 \tag{2.5}$$

for all functions $f \in \mathcal{F}(p)$.

We prove Lemma 2.1 by applying a general technique for constructing equations such as (2.5) from discrete probability mass functions which is of independent interest; see [19]. We begin with Proposition 2.1 below, a discrete version of the density approach to the Stein equation.

Proposition 2.1. *Let Z have probability mass function p with support the integer interval I , and let $\psi(k)$ be given by (2.4) for $k \in I$. If a random variable X with support I has mass function p then, for all $f \in \mathcal{F}(p)$,*

$$\mathbb{E}[\Delta f(X - 1) + \psi(X)f(X)] = 0. \tag{2.6}$$

Conversely, if (2.6) holds for all $f(k) = \mathbf{1}(k = \ell)$, $\ell \in I$, then X has mass function p .

Remark 2.1. The functions $f(k) = \mathbf{1}(k = \ell)$, $\ell \in I$, are in $\mathcal{F}(p)$, so Proposition 2.1 implies that a random variable X with support I has mass function p if and only if (2.6) holds for all $f \in \mathcal{F}(p)$.

The statement in Proposition 2.1 is equivalent to Theorem 1.1 given in [19] under a different assumption, namely that equality (2.6) holds with g replacing f for all functions for which $\sum_{k \in I} \Delta(g(k)p(k)) = 0$. We note that their setup would translate to test functions $f(k) = g(k + 1)$.

Proof. Let p be a real-valued function defined on the integer interval $[a, b + 1] \cap \mathbb{Z}$ for $a, b \in \mathbb{Z}$ with $a \leq b$. Applying the summation by parts formula in the first line below, we obtain

$$\sum_{k=a}^b f(k)\Delta p(k) = -\sum_{k=a}^b p(k + 1)\Delta f(k) + f(b + 1)p(b + 1) - f(a)p(a) \tag{2.7}$$

$$= -\sum_{k=a+1}^{b+1} p(k)\Delta f(k - 1) + f(b + 1)p(b + 1) - f(a)p(a)$$

$$= -\sum_{k=a}^b p(k)\Delta f(k - 1) + f(b)p(b + 1) - f(a - 1)p(a). \tag{2.8}$$

If $p(b + 1) = 0$ and $f(a - 1) = 0$, then we obtain

$$\sum_{k=a}^b f(k)\Delta p(k) = -\sum_{k=a}^b p(k)\Delta f(k - 1). \tag{2.9}$$

Hence, (2.6) holds when p is a probability mass function with support $[a, b] \cap \mathbb{Z}$ and $f(a - 1) = 0$.

The case where I is an infinite integer interval follows by applying Abel’s lemma on summation by parts as modified by [9]. In particular, if either (2.8) or the left-hand side of (2.7) is convergent upon replacing b by ∞ , and if $\lim_{n \rightarrow \infty} f(n)p(n + 1) = 0$, then Equation (3a) of [9] shows that (2.9) holds upon replacing b by ∞ , completing the argument when $I = [a, \infty)$. Similarly, Equation (3b) of [9] can be used to argue the case when $I = \mathbb{Z}$.

For $I = (-\infty, b]$, (2.8) gives

$$\begin{aligned} \mathbb{E}\psi(X)f(X) &= \sum_{k=-\infty}^b f(k)\Delta p(k) \\ &= \lim_{a \rightarrow -\infty} \sum_{k=a}^b f(k)\Delta p(k) \\ &= \lim_{a \rightarrow -\infty} \left(-\sum_{k=a}^b p(k)\Delta f(k - 1) + f(b)p(b + 1) - f(a - 1)p(a) \right). \end{aligned}$$

Since $p(b + 1) = 0$ and $p(a)f(a - 1) \rightarrow 0$ as $a \rightarrow -\infty$, we obtain

$$\mathbb{E}\psi(X)f(X) = -\lim_{a \rightarrow -\infty} \sum_{k=a}^b p(k)\Delta f(k - 1) = -\mathbb{E}\Delta f(X - 1)$$

and (2.6) holds.

Conversely, if X with support I satisfies (2.6) for all functions $f(k) = \mathbf{1}(k = \ell)$ for $\ell \in I$, then

$$\begin{aligned} 0 &= \mathbb{E}[\Delta f(X - 1) + \psi(X)f(X)] \\ &= \sum_{k \in I} \mathbb{P}(X = k) \{ (f(k) - f(k - 1)) + \psi(k)f(k) \} \\ &= \mathbb{P}(X = \ell) - \mathbb{P}(X = \ell + 1) + \mathbb{P}(X = \ell) \frac{\mathbb{P}(Z = \ell + 1) - \mathbb{P}(Z = \ell)}{\mathbb{P}(Z = \ell)}, \end{aligned}$$

and rearranging gives

$$\frac{\mathbb{P}(X = \ell + 1)}{\mathbb{P}(X = \ell)} = \frac{\mathbb{P}(Z = \ell + 1)}{\mathbb{P}(Z = \ell)}.$$

Hence, if I is $[a, b]$ or $[a, \infty)$, we obtain, for all $j \in I$,

$$\frac{\mathbb{P}(X = j)}{\mathbb{P}(X = a)} = \prod_{\ell=a}^{j-1} \frac{\mathbb{P}(X = \ell + 1)}{\mathbb{P}(X = \ell)} = \prod_{\ell=a}^{j-1} \frac{\mathbb{P}(Z = \ell + 1)}{\mathbb{P}(Z = \ell)} = \frac{\mathbb{P}(Z = j)}{\mathbb{P}(Z = a)}.$$

Summing over $j \in I$ yields $\mathbb{P}(X = a) = \mathbb{P}(Z = a)$, whence $\mathbb{P}(X = j) = \mathbb{P}(Z = j)$ for all $j \in I$. Similarly, we may handle the remaining case where $I = (-\infty, b]$.

Given a characterization produced by Proposition 2.1, the following corollary produces varieties of characterizations for the same distribution, each corresponding to a choice of a function c possessing certain mild properties.

Corollary 2.1. *Let Z be a random variable with probability mass function p and support $I = [a, b] \cap \mathbb{Z}$, where $a \leq b$, $a, b \in \mathbb{Z}$, and let ψ be given by (2.4). Then, for all functions $c: [a - 1, b] \cap \mathbb{Z} \rightarrow \mathbb{R} \setminus \{0\}$, a random variable X with support I has mass function p if and only if*

$$\mathbb{E}[c(X - 1)\Delta f(X - 1) + [c(X)\psi(X) + c(X) - c(X - 1)]f(X)] = 0 \tag{2.10}$$

for all functions $f \in \mathcal{F}(p)$.

Proof. Suppose that (2.10) holds for X for all functions in $f \in \mathcal{F}(p)$. Since $c(\ell) \neq 0$ for all $\ell \in I$, the functions $f(k) = c(\ell)^{-1}\mathbf{1}(k = \ell)$, $\ell \in I$, all lie in $\mathcal{F}(p)$, so X has mass function p by Proposition 2.1. For the converse, note that $f \in \mathcal{F}(p)$ implies that $cf \in \mathcal{F}(p)$. Hence, if X has mass function p then (2.10) follows from Proposition 2.1 upon replacing $f(x)$ by $c(x)f(x)$.

Proposition 2.1 applied to Z with mass function p of the Poisson distribution $\mathcal{P}(\lambda)$, having $\psi(k) = \lambda/(k + 1) - 1$ by (2.4) for all $k \in \mathbb{N}_0$, yields, for all functions $f \in \mathcal{F}(p)$,

$$\mathbb{E}\Delta f(Z - 1) = \mathbb{E}\left[1 - \frac{\lambda}{Z + 1}\right]f(Z), \quad \text{that is,} \quad \mathbb{E}f(Z - 1) = \mathbb{E}\left[\frac{\lambda}{Z + 1}\right]f(Z),$$

a nonstandard version of a characterization of the Poisson. An extension of Corollary 2.1 to the case of infinite support produces the usual characterization by the choice $c(k) = k + 1$ and the substitution $g(k) = f(k - 1)$. Naturally, additional characterizations are produced when using different choices of c .

Remark 2.2. When $I = [a, b] \cap \mathbb{Z}$, we automatically have $\psi(b) = -1$, in which case $c(b)$ does not appear in (2.10), and may be assigned a value arbitrarily.

We now apply Corollary 2.1 to the distribution of $S_n^{\alpha, \beta, m}$, the number of white balls drawn from the Pólya–Eggenberger urn by time n . We suppress α, β , and m for notational ease unless clarity demands it. It is well known, and not difficult to verify, that the distribution $p_k = \mathbb{P}(S_n = k)$, $k \in \mathbb{Z}$, satisfies

$$p_k = \binom{n}{k} \frac{(\alpha/m)_k (\beta/m)_{n-k}}{(\alpha/m + \beta/m)_n}, \tag{2.11}$$

where $(x)_0 = 1$ and otherwise $(x)_k = x(x + 1) \cdots (x + k - 1)$ is the rising factorial. Distribution (2.11) is also known as the beta-binomial and the negative hypergeometric distribution; see [35]. We now have the ingredients to prove Lemma 2.1.

Proof of Lemma 2.1. Taking differences in (2.11) for $k = 0, \dots, n - 1$ yields

$$\Delta p_k = \binom{n}{k} \frac{(\alpha/m)_k (\beta/m)_{n-k}}{(\alpha/m + \beta/m)_n} \left\{ \frac{(n - k)(\alpha/m + k)}{(k + 1)(\beta/m + n - k - 1)} - 1 \right\},$$

while, for $k = n$,

$$\Delta p_n = -p_n.$$

Hence, with $\psi(k) = \Delta p_k / p_k$ as in (2.4) we obtain, for $k = 0, \dots, n - 1$,

$$\psi(k) = \frac{(n - k)(\alpha/m + k) - (k + 1)(\beta/m + n - k - 1)}{(k + 1)(\beta/m + n - k - 1)}$$

and $\psi(n) = -1$.

In applying Corollary 2.1, as $\psi(n) = -1$, we may take the value $c(n)$ arbitrarily; see Remark 2.2. In particular, taking $c(k) = (k + 1)(\beta/m + n - k - 1)$ for all $k = 0, \dots, n - 1$ and $c(n) = n$ we obtain (2.5).

The next lemma is instrumental in calculating the higher moments of $S_n^{\alpha, \beta, m}$. We let $[x]_0 = 1$, and otherwise set $[x]_k = x(x - 1) \cdots (x - k + 1)$, the falling factorial.

Lemma 2.2. *For all nonnegative integers n, a , and b , we have*

$$\mathbb{E}[[S_n^{\alpha, \beta, m}]_a [n - S_n^{\alpha, \beta, m}]_b] = \frac{[n]_{a+b} (\alpha/m)_a (\beta/m)_b}{(\alpha/m + \beta/m)_{a+b}}. \tag{2.12}$$

Proof. First we note that both sides of (2.12) are 0 when $a + b \geq n + 1$. This is clear for the right-hand side, as the falling factorial $[n]_{a+b}$ is 0. For the left-hand side, if $S_n \leq a - 1$ then $[S_n]_a = 0$. On the other hand, if $S_n \geq a$ then $b - 1 \geq n - a \geq n - S_n$, in which case $[n - S_n]_b$ is 0.

Now assume that $n \geq a + b$. For any $k = 0, 1, \dots, n$, we have

$$\begin{aligned} & [k]_a [n - k]_b \mathbb{P}(S_n^{\alpha, \beta, m} = k) \\ &= [k]_a [n - k]_b \binom{n}{k} \frac{(\alpha/m)_k (\beta/m)_{n-k}}{(\alpha/m + \beta/m)_n} \\ &= \frac{[n]_{a+b} (\alpha/m)_a (\beta/m)_b}{(\alpha/m + \beta/m)_{a+b}} \binom{n - a - b}{k - a} \frac{(\alpha/m + a)_{k-a} (\beta/m + b)_{n-k-b}}{(\alpha/m + \beta/m + a + b)_{n-a-b}} \\ &= \frac{[n]_{a+b} (\alpha/m)_a (\beta/m)_b}{(\alpha/m + \beta/m)_{a+b}} \mathbb{P}(a + S_{n-a-b}^{\alpha+am, \beta+bm, m} = k). \end{aligned}$$

Summing over $k = 0, 1, \dots, n$ and using the fact that the support of S_r is $\{0, \dots, r\}$ yields (2.12).

If Z has the limiting beta distribution $\mathcal{B}(\alpha/m, \beta/m)$ with density (1.2), using (2.12), we obtain

$$\begin{aligned} \mathbb{E}\left[\frac{[S_n]_a [n - S_n]_b}{n^{a+b}}\right] &= \frac{[n]_{a+b} (\alpha/m)_a (\beta/m)_b}{n^{a+b} (\alpha/m + \beta/m)_{a+b}} \\ &= \frac{[n]_{a+b}}{n^{a+b}} \frac{B(\alpha/m + a, \beta/m + b)}{B(\alpha/m, \beta/m)} \\ &= \frac{[n]_{a+b}}{n^{a+b}} \mathbb{E}[Z^a (1 - Z)^b], \end{aligned} \tag{2.13}$$

that is, the scaled falling factorial moments of S_n and the power moments of Z differ only by factors of order $1/n$. This observation can be used to provide a proof of convergence in distribution of $W_n = S_n/n$ to Z by the method of moments, but without a bound on the distributional distance.

3. Bounds for the Pólya–Eggenberger urn model

Theorem 1.1 provides an explicit bound in Wasserstein distance of order $O(1/n)$ between the distribution of W_n , the fraction of white balls drawn from the urn by time n , and the limiting beta distribution. For approximating a discrete distribution by a continuous distribution, the

Wasserstein distance d_W is a typical distance to use; see, for example, [17]. For random variables X and Y , this distance is given by

$$d_W(X, Y) = \sup_{h \in \text{Lip}(1)} |\mathbb{E}[h(X)] - \mathbb{E}[h(Y)]|, \tag{3.1}$$

where $\text{Lip}(1) = \{h: |h(x) - h(y)| \leq |x - y|\}$, the class of all Lipschitz-continuous functions with Lipschitz constant less than or equal to 1. The Wasserstein distance defines a metric on the set of probability measures on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$, the set of real numbers equipped with their Borel σ -field. On this space, convergence under the Wasserstein distance implies weak convergence. On $([0, 1], \mathcal{B}([0, 1]))$, the Wasserstein distance metrizes weak convergence; see [17].

Remark 3.1. The function $h(x) = x(1 - x)\mathbf{1}_{\{x \in [0,1]\}}$ is in $\text{Lip}(1)$, and applying (2.13) with $a = b = 1$ we obtain, for all $\alpha \geq 1, \beta \geq 1$, and $m \geq 1$,

$$\begin{aligned} d_W(W_n, Z) &\geq |\mathbb{E}[h(W_n)] - \mathbb{E}h(Z)| \\ &= \left| \left(\frac{\lfloor n \rfloor_2}{n^2} - 1 \right) \mathbb{E}[Z(1 - Z)] \right| \\ &= \frac{1}{n} \frac{\alpha\beta}{(\alpha + \beta + m)(\alpha + \beta)}. \end{aligned}$$

Thus, the $1/n$ order of the bound in Theorem 1.1 cannot be improved.

Remark 3.2. Theorem 4.3 of [12] provides a bound of order $1/n$ for the beta approximation to the Pólya–Eggenberger urn for test functions with bounded first and second derivatives using an exchangeable pair coupling. The results in [12] differ from ours in two significant ways. Firstly, the bound in Theorem 4.3 of [12] is expressed in terms of two nonexplicit constants C_1 and C_2 that are defined in Proposition 3.8 of [12]. Lemma 3.4 below provides values of C_1 . The lack of an explicit expression for C_2 in [12] can be explained by the fact that the solution there is given in terms of ratios of functions which are related to incomplete beta functions, for which a uniform bound would be difficult.

A more important difference between the present work and [12] is that expressing the bound of the latter, presently given in terms of twice differentiable functions, in terms of a bound in a metric, say d_2 , obtained from twice differentiable functions in the same way that Lipschitz functions yield the Wasserstein metric d_W , we have $d_2 \leq d_W$ with equality everywhere not holding. Hence, Theorem 1.1 implies bounds in the d_2 metric, while the reverse does not hold.

In the following we set our test functions h to be 0 outside the unit interval $[0, 1]$. For $y > 0$, set

$$\Delta_y f(x) = f(x + y) - f(x),$$

and, for a real-valued function g on $[0, 1]$, we let $\|g\| = \sup_{w \in [0,1]} |g(w)|$, the supremum norm of g . In the following we recall, with the help of Rademacher's theorem, that a function h is in $\text{Lip}(1)$ if and only if it is absolutely continuous with respect to the Lebesgue measure with an almost everywhere derivative bounded in absolute value by 1.

Lemma 3.1 below shows that, for all $\{\alpha, \beta\} \subset (0, \infty)$ and functions h for which the expectation $\mathcal{B}_{\alpha,\beta}h$ exists,

$$f_{\alpha,\beta}(w) = \frac{1}{w^\alpha(1 - w)^\beta} \int_0^w u^{\alpha-1}(1 - u)^{\beta-1}(h(u) - \mathcal{B}_{\alpha,\beta}h) du, \quad w \in [0, 1], \tag{3.2}$$

is the unique bounded solution of the Stein equation (2.3). We continue to omit subscripts when the context makes it clear which parameters are used. In the proof below we invoke Lemma 3.4 which yields bounds on the supremum norm of the derivative f' in terms of that same norm on the derivative h' of the test function.

Proof of Theorem 1.1. For h a given function in $\text{Lip}(1)$, let $f = f_{\alpha/m, \beta/m}$ be the solution of the Stein equation (2.3) given in (3.2). Replacing $f(z)$ by $f(z/n)$ and dividing by n in (2.5) results in

$$\begin{aligned} 0 &= \mathbb{E}S_n \left(\frac{\beta}{nm} + 1 - W_n \right) \Delta_{1/n} f \left(W_n - \frac{1}{n} \right) \\ &\quad + \mathbb{E} \left[(n - S_n) \left(\frac{\alpha}{nm} + W_n \right) - S_n \left(\frac{\beta}{nm} + 1 - W_n \right) \right] f(W_n) \\ &= \mathbb{E} \left[nW_n \left(\frac{\beta}{nm} + 1 - W_n \right) \Delta_{1/n} f \left(W_n - \frac{1}{n} \right) + \left\{ \frac{\alpha}{m} (1 - W_n) - \frac{\beta}{m} W_n \right\} f(W_n) \right]. \end{aligned}$$

Applying this identity in the Stein equation (2.3), with α and β replaced by α/m and β/m , respectively, and invoking Lemma 3.4 below to yield the existence and boundedness of f' , we obtain

$$\begin{aligned} \mathbb{E}h(W_n) - \mathcal{B}h &= \mathbb{E} \left[W_n(1 - W_n)f'(W_n) + \left[\frac{\alpha}{m}(1 - W_n) - \frac{\beta}{m}W_n \right] f(W_n) \right] \\ &= \mathbb{E} \left[W_n(1 - W_n)f'(W_n) - nW_n \left(\frac{\beta}{nm} + 1 - W_n \right) \Delta_{1/n} f \left(W_n - \frac{1}{n} \right) \right] \\ &= \mathbb{E} \left[W_n(1 - W_n)f'(W_n) - nW_n(1 - W_n) \Delta_{1/n} f \left(W_n - \frac{1}{n} \right) \right] + R_1, \end{aligned} \tag{3.3}$$

where, using Lemma 2.2 to calculate moments, we obtain

$$|R_1| = \frac{\beta}{m} \left| \mathbb{E}W_n \Delta_{1/n} f \left(W_n - \frac{1}{n} \right) \right| \leq \frac{\beta}{nm} \|f'\| \mathbb{E}W_n = \frac{\alpha\beta}{nm(\alpha + \beta)} \|f'\|. \tag{3.4}$$

Writing the difference in (3.3) as an integral, we have

$$\begin{aligned} &\mathbb{E} \left[W_n(1 - W_n)f'(W_n) - nW_n(1 - W_n) \Delta_{1/n} f \left(W_n - \frac{1}{n} \right) \right] \\ &= \mathbb{E}W_n(1 - W_n) \left(f'(W_n) - n \int_{W_n-1/n}^{W_n} f'(x) dx \right) \\ &= n \mathbb{E} \int_{W_n-1/n}^{W_n} W_n(1 - W_n)(f'(W_n) - f'(x)) dx \\ &= n \mathbb{E} \int_{W_n-1/n}^{W_n} (W_n(1 - W_n)f'(W_n) - x(1 - x)f'(x)) dx - R_2, \end{aligned} \tag{3.5}$$

where

$$R_2 = n \mathbb{E} \int_{W_n-1/n}^{W_n} (W_n(1 - W_n) - x(1 - x))f'(x) dx.$$

For R_2 , using the fact that the solution f of the Stein equation equals 0 for $x \notin [0, 1]$ to obtain the first inequality, we obtain

$$\begin{aligned}
 |R_2| &= \left| n\mathbb{E} \int_{W_n-1/n}^{W_n} f'(x) \int_x^{W_n} (1-2y) \, dy \, dx \right| \\
 &\leq \|f'\| n\mathbb{E} \int_{W_n-1/n}^{W_n} \int_x^{W_n} dy \, dx \\
 &= \|f'\| n\mathbb{E} \int_{W_n-1/n}^{W_n} (W_n - x) \, dx \\
 &= \frac{1}{2n} \|f'\|. \tag{3.6}
 \end{aligned}$$

For the first term in (3.5), substituting using the Stein equation (2.3) with α and β replaced by α/m and β/m , respectively, we obtain

$$\begin{aligned}
 &n\mathbb{E} \int_{W_n-1/n}^{W_n} (W_n(1 - W_n)f'(W_n) - x(1 - x)f'(x)) \, dx \\
 &= n\mathbb{E} \int_{W_n-1/n}^{W_n} (h(W_n) - h(x)) \, dx \\
 &\quad + n\mathbb{E} \int_{W_n-1/n}^{W_n} \frac{1}{m} [(\beta W_n - \alpha(1 - W_n))f(W_n) - (\beta x - \alpha(1 - x))f(x)] \, dx \\
 &= n\mathbb{E} \int_{W_n-1/n}^{W_n} \left(\int_x^{W_n} h'(y) \, dy + \frac{1}{m} \int_x^{W_n} [(\beta y - \alpha(1 - y))f(y)]' \, dy \right) \, dx \\
 &= n\mathbb{E} \int_{W_n-1/n}^{W_n} \left(\int_x^{W_n} h'(y) \, dy \right. \\
 &\quad \left. + \frac{1}{m} \int_x^{W_n} [(\beta y - \alpha(1 - y))f'(y) + (\beta + \alpha)f(y)] \, dy \right) \, dx.
 \end{aligned}$$

We bound the two integrals separately. Firstly,

$$\left| n\mathbb{E} \int_{W_n-1/n}^{W_n} (h(W_n) - h(x)) \, dx \right| \leq n \|h'\| \mathbb{E} \int_{W_n-1/n}^{W_n} (W_n - x) \, dx = \frac{1}{2n} \|h'\|. \tag{3.7}$$

Next, recalling that $0 \leq W_n \leq 1$ and noting that $|(\beta y - \alpha(1 - y))| \leq \alpha \vee \beta$ for $0 \leq y \leq 1$,

$$\begin{aligned}
 &\left| \frac{n}{m} \mathbb{E} \int_{W_n-1/n}^{W_n} \int_x^{W_n} (\beta y - \alpha(1 - y))f'(y) \, dy \, dx \right| \\
 &= \left| \frac{n}{m} \mathbb{E} \int_{W_n-1/n}^{W_n} (\beta y - \alpha(1 - y))f'(y) \int_{W_n-1/n}^y dx \, dy \right| \\
 &\leq \frac{n}{m} \|f'\| (\alpha \vee \beta) \mathbb{E} \int_{W_n-1/n}^{W_n} \int_{W_n-1/n}^y dx \, dy \\
 &= \frac{1}{2nm} \|f'\| (\alpha \vee \beta). \tag{3.8}
 \end{aligned}$$

Arguing in a similar fashion, we obtain

$$\begin{aligned} \left| \frac{n}{m} \mathbb{E} \int_{W_{n-1/n}}^{W_n} \int_x^{W_n} (\beta + \alpha) f(y) dy dx \right| &\leq \frac{n}{m} \|f\| (\beta + \alpha) \mathbb{E} \int_{W_{n-1/n}}^{W_n} \int_{W_{n-1/n}}^y dx dy \\ &= \frac{1}{2nm} \|f\| (\alpha + \beta). \end{aligned} \tag{3.9}$$

Collecting the bounds (3.4), (3.6), (3.7), (3.8), and (3.9) yields

$$\begin{aligned} &|\mathbb{E}h(W_n) - \mathbb{E}h(W)| \\ &\leq \left(\frac{m + \alpha \vee \beta}{2nm} + \frac{\alpha\beta}{nm(\alpha + \beta)} \right) \|f'\| + \frac{1}{2n} \|h'\| + \frac{1}{2nm} (\alpha + \beta) \|f\|. \end{aligned}$$

Note that, by Lemma 3.2,

$$\frac{1}{2nm} (\alpha + \beta) \|f\| \leq \frac{1}{2nm} \left(\frac{2}{\alpha/m + \beta/m} \right) \|h'\| = \frac{1}{n} \|h'\|.$$

The theorem now follows by invoking Lemma 3.4.

Lemma 3.1. *For any $\{\alpha, \beta\} \subset (0, \infty)$ and real-valued function h on $[0, 1]$ such that the expectation $\mathcal{B}_{\alpha,\beta}h$ of h exists, the function f given by (3.2) is the unique bounded solution of (2.3).*

Proof. It is straightforward to verify that f as given in (3.2) is a solution of (2.3). Writing the associated homogeneous equation as

$$(w^{\alpha-1}(1-w)^{\beta-1})^{-1} (w^\alpha(1-w)^\beta g(w))' = 0,$$

we find that all solutions to (2.3) are given by

$$f(w) + cg(w) \quad \text{for some } c \in \mathbb{R}, \quad \text{where } g(w) = \frac{1}{w^\alpha(1-w)^\beta}.$$

The claim follows since $g(w)$ is unbounded at the endpoints of the unit interval for all $c \neq 0$, and Lemma 3.4 below demonstrates that $f(w)$ is bounded.

Since the expectation of $h(Z) - \mathcal{B}h$ is 0 when $Z \sim \mathcal{B}(\alpha, \beta)$, we may also write

$$f(w) = -\frac{1}{w^\alpha(1-w)^\beta} \int_w^1 u^{\alpha-1}(1-u)^{\beta-1} (h(u) - \mathcal{B}h) du. \tag{3.10}$$

From Proposition 3.8 of [12], we quote the following result.

Lemma 3.2. *The solution f , given in (3.2), of (2.3) for h a Lipschitz function on $[0, 1]$ satisfies*

$$\|f\| \leq \frac{2}{\alpha + \beta} \|h'\|.$$

The cases in the bounds of Lemma 3.4 reflect the behavior of the function

$$g(w) = w^{\alpha-1}(1-w)^{\beta-1} \quad \text{for } w \in [0, 1], \tag{3.11}$$

as described in Lemma 3.3. In the following we will use the terms decreasing and increasing in the nonstrict manner, for example, a constant function is both increasing and decreasing. Let

$$x_{\alpha,\beta} = \frac{\alpha - 1}{\alpha + \beta - 2}.$$

TABLE 1.

	$\beta < 1$	$\beta = 1$	$\beta > 1$
$\alpha < 1$	Decreasing on $[0, x_{\alpha, \beta}]$, Increasing on $[x_{\alpha, \beta}, 1]$	Decreasing	Decreasing
$\alpha = 1$	Increasing	Constant	Decreasing
$\alpha > 1$	Increasing	Increasing	Increasing on $[0, x_{\alpha, \beta}]$, Decreasing on $[x_{\alpha, \beta}, 1]$

Lemma 3.3. For $\{\alpha, \beta\} \subset (-1, \infty)$, the function $g: [0, 1] \rightarrow [0, \infty)$ given in (3.11) has the behavior outlined in Table 1.

Proof. Clearly, when $\alpha = 1$ and $\beta = 1$, the function $g(w)$ is constant. Otherwise, taking the derivative in (3.11) yields

$$g'(w) = (\alpha - 1)w^{\alpha-2}(1 - w)^{\beta-1} - (\beta - 1)w^{\alpha-1}(1 - w)^{\beta-2} \\ = w^{\alpha-2}(1 - w)^{\beta-2}\{(\alpha - 1)(1 - w) - (\beta - 1)w\}.$$

The expression is nonnegative if and only if

$$(\alpha - 1)(1 - w) \geq (\beta - 1)w. \tag{3.12}$$

When $\alpha \geq 1$ and $\beta \leq 1$, inequality (3.12) is always satisfied. Similarly, (3.12) holds with the nonstrict inequality reversed when $\alpha \leq 1$ and $\beta \geq 1$. The remaining two cases $\alpha < 1, \beta < 1$ and $\alpha > 1, \beta > 1$ follow by solving the inequality.

Our next result bounds the magnitude of the derivative of the solution f in terms of h .

Lemma 3.4. For $\{\alpha, \beta\} \subset (0, \infty)$, let $f = f_{\alpha, \beta, h}$ be the solution to (2.3) given by (3.2) for an absolutely continuous function h . Then

$$\|f'\| \leq b_0\|h - \mathcal{B}h\| + b_1\|h'\| \leq (b_0 + b_1)\|h'\|, \tag{3.13}$$

where $b_0 = b_0(\alpha, \beta)$ and $b_1 = b_1(\alpha, \beta)$ are given by

$$b_0 = \begin{cases} 4 \max(|\alpha - 1|; |\beta - 1|) & \text{if } \alpha \leq 2, \beta \leq 2, \\ (\alpha + \beta - 2)^2 \max\left(\frac{\alpha - 1}{(\alpha - 2)^2}; \frac{|\beta - 1|}{\beta^2}\right) & \text{if } \alpha > 2, \beta \leq 2, \\ (\alpha + \beta - 2)^2 \max\left(\frac{|\alpha - 1|}{\alpha^2}; \frac{\beta - 1}{(\beta - 2)^2}\right) & \text{if } \alpha \leq 2, \beta > 2, \\ (\alpha + \beta - 2)^2 \max\left(\frac{1}{\alpha - 1}; \frac{1}{\beta - 1}\right) & \text{if } \alpha > 2, \beta > 2, \end{cases}$$

and

$$b_1 = \begin{cases} 4\left(1 + \frac{\max(\alpha; \beta)}{\alpha + \beta}\right) & \text{if } \alpha \leq 2, \beta \leq 2, \\ \frac{(\alpha + \beta - 2)^2}{\min(\alpha - 2, \beta)^2} + 2 \max\left(\frac{\alpha}{\alpha - 2}; 1\right) & \text{if } \alpha > 2, \beta \leq 2, \\ \frac{(\alpha + \beta - 2)^2}{\min(\alpha, \beta - 2)^2} + 2 \max\left(1; \frac{\beta}{\beta - 2}\right) & \text{if } \alpha \leq 2, \beta > 2, \\ \frac{(\alpha + \beta - 2)^2}{\min(\alpha - 1, \beta - 1)^2} + 2 \max\left(\frac{\alpha}{\alpha - 1}; \frac{\beta}{\beta - 1}\right) & \text{if } \alpha > 2, \beta > 2. \end{cases}$$

Proof. By replacing h by $h - \mathcal{B}_{\alpha,\beta}h$ we may assume that $\mathcal{B}_{\alpha,\beta}h = 0$. Rewriting the Stein equation (2.3) yields

$$w(1 - w)f'(w) = h(w) + (\beta w - \alpha(1 - w))f(w),$$

so to show (3.13) it suffices to demonstrate that, for all $w \in [0, 1]$,

$$|h(w) + (\beta w - \alpha(1 - w))f(w)| \leq (b_0\|h\| + b_1\|h'\|)w(1 - w).$$

Using (3.2) and integration by parts, we obtain

$$\begin{aligned} \alpha(1 - w)f(w) &= h(w) + \frac{\beta - 1}{w^\alpha(1 - w)^{\beta-1}} \int_0^w u^\alpha(1 - u)^{\beta-2}h(u) \, du \\ &\quad - \frac{1}{w^\alpha(1 - w)^{\beta-1}} \int_0^w u^\alpha(1 - u)^{\beta-1}h'(u) \, du \end{aligned}$$

and, now applying (3.10),

$$\begin{aligned} \beta wf(w) &= -h(w) - \frac{\alpha - 1}{w^{\alpha-1}(1 - w)^\beta} \int_w^1 u^{\alpha-2}(1 - u)^\beta h(u) \, du \\ &\quad - \frac{1}{w^{\alpha-1}(1 - w)^\beta} \int_w^1 u^{\alpha-1}(1 - u)^\beta h'(u) \, du. \end{aligned}$$

Hence,

$$\begin{aligned} &h(w) + (\beta w - \alpha(1 - w))f(w) \\ &= \frac{1}{w^\alpha(1 - w)^{\beta-1}} \int_0^w u^\alpha(1 - u)^{\beta-2}(-(\beta - 1)h(u) + (1 - u)h'(u)) \, du + \beta wf(w) \\ &= \frac{1}{w^{\alpha-1}(1 - w)^\beta} \int_w^1 u^{\alpha-2}(1 - u)^\beta(-(\alpha - 1)h(u) - uh'(u)) \, du \\ &\quad - \alpha(1 - w)f(w). \end{aligned}$$

From Lemma 3.2 we immediately have the bounds

$$|\beta wf(w)| \leq \frac{2\beta}{\alpha + \beta}w\|h'\| \quad \text{and} \quad |\alpha(1 - w)f(w)| \leq \frac{2\alpha}{\alpha + \beta}(1 - w)\|h'\|. \tag{3.14}$$

As $0 \leq u \leq 1$,

$$\left| \frac{1}{w^\alpha(1-w)^{\beta-1}} \int_0^w u^\alpha(1-u)^{\beta-2}(-(\beta-1)h(u) + (1-u)h'(u)) du \right| \leq (|\beta-1|\|h\| + \|h'\|) \frac{1}{w^\alpha(1-w)^{\beta-1}} \int_0^w u^\alpha(1-u)^{\beta-2} du.$$

When $u^\alpha(1-u)^{\beta-2}$ is increasing on $[0, x_*]$ then, for $w \in [0, x_*]$, we can bound this expression by

$$\|h\|w(1-w) \frac{|\beta-1|}{(1-x_*)^2} + \|h'\|w(1-w) \frac{1}{(1-x_*)^2},$$

and now using the first inequality in (3.14), we obtain

$$\begin{aligned} & |h(w) + (\beta w - \alpha(1-w))f(w)| \\ & \leq \|h\|w(1-w) \frac{|\beta-1|}{(1-x_*)^2} + \|h'\|w(1-w) \left(\frac{1}{(1-x_*)^2} + \frac{2\beta}{(\alpha+\beta)(1-x_*)} \right) \\ & \leq \frac{|\beta-1|}{(1-x_*)^2} \|h\|w(1-w) + \left(\frac{1}{(1-x_*)^2} + \frac{2\beta}{(\alpha+\beta)(1-x_*)} \right) \|h'\|w(1-w). \end{aligned}$$

Similarly,

$$\left| \frac{1}{w^{\alpha-1}(1-w)^\beta} \int_w^1 u^{\alpha-2}(1-u)^\beta(-(\alpha-1)h(u) - uh'(u)) du \right| \leq (|\alpha-1|\|h\| + \|h'\|) \frac{1}{w^{\alpha-1}(1-w)^\beta} \int_w^1 u^{\alpha-2}(1-u)^\beta du,$$

and if $u^{\alpha-2}(1-u)^\beta$ is decreasing on $[x_*, 1]$ then, for $w \in [x_*, 1]$, we can bound this expression by

$$\|h\|w(1-w) \frac{|\alpha-1|}{x_*^2} + \|h'\|w(1-w) \frac{1}{x_*^2}.$$

Now using the second inequality in (3.14), we obtain

$$\begin{aligned} & |h(w) + (\beta w - \alpha(1-w))f(w)| \\ & \leq \|h\|w(1-w) \frac{|\alpha-1|}{x_*^2} + \left(\frac{1}{x_*^2} + \frac{2\alpha}{(\alpha+\beta)x_*} \right) \|h'\|w(1-w). \end{aligned}$$

Hence, we may take

$$b_0 = \max\left(\frac{|\alpha-1|}{x_*^2}, \frac{|\beta-1|}{(1-x_*)^2} \right) \tag{3.15}$$

and

$$b_1 = \max\left\{ \frac{1}{x_*^2} + \frac{2\alpha}{(\alpha+\beta)x_*}; \frac{1}{(1-x_*)^2} + \frac{2\beta}{(\alpha+\beta)(1-x_*)} \right\}. \tag{3.16}$$

In view of Lemma 3.3 we distinguish four cases.

Case 1: $\alpha \leq 2$ and $\beta \leq 2$. By Lemma 3.3, $u^\alpha(1-u)^{\beta-2}$ is increasing and $u^{\alpha-2}(1-u)^\beta$ is decreasing. Setting $x_* = \frac{1}{2}$, by (3.15) and (3.16), we obtain

$$b_0 = 4 \max(|\alpha-1|; |\beta-1|) \quad \text{and} \quad b_1 = 4 \left(1 + \frac{\max(\alpha; \beta)}{\alpha + \beta} \right).$$

Case 2: $\alpha > 2$ and $\beta \leq 2$. In this case, from Lemma 3.3, $u^\alpha(1-u)^{\beta-2}$ is increasing and $u^{\alpha-2}(1-u)^\beta$ is decreasing on $[x_{\alpha-1,\beta+1}, 1]$. Setting $x_* = x_{\alpha-1,\beta+1}$ and noting that $x_{\alpha,\beta} + x_{\beta,\alpha} = 1$, by (3.15) and (3.16), we obtain

$$b_0 = \max\left(\frac{\alpha - 1}{x_{\alpha-1,\beta+1}^2}; \frac{|\beta - 1|}{x_{\beta+1,\alpha-1}^2}\right) = (\alpha + \beta - 2)^2 \max\left(\frac{\alpha - 1}{(\alpha - 2)^2}; \frac{|\beta - 1|}{\beta^2}\right)$$

and

$$b_1 = \max\left\{\frac{1}{x_{\alpha-1,\beta+1}^2} + \frac{2\alpha}{(\alpha + \beta)x_{\alpha-1,\beta+1}}; \frac{1}{x_{\beta+1,\alpha-1}^2} + \frac{2\beta}{(\alpha + \beta)x_{\beta+1,\alpha-1}}\right\} \\ \leq \frac{(\alpha + \beta - 2)^2}{\min(\alpha - 2, \beta)^2} + 2\left(\frac{\alpha + \beta - 2}{\alpha + \beta}\right) \max\left(\frac{\alpha}{\alpha - 2}; 1\right),$$

and bounding $(\alpha + \beta - 2)/(\alpha + \beta)$ by 1 gives the assertion.

Case 3: $\alpha \leq 2$ and $\beta > 2$. In this case, from Lemma 3.3, $u^\alpha(1-u)^{\beta-2}$ is increasing on $[0, x_{\alpha+1,\beta-1}]$ and $u^{\alpha-2}(1-u)^\beta$ is decreasing. Setting $x_* = x_{\alpha+1,\beta-1}$, by (3.15) and (3.16), we obtain

$$b_0 = \max\left(\frac{|\alpha - 1|}{x_{\alpha+1,\beta-1}^2}; \frac{\beta - 1}{x_{\beta-1,\alpha+1}^2}\right) = (\alpha + \beta - 2)^2 \max\left(\frac{|\alpha - 1|}{\alpha^2}; \frac{\beta - 1}{(\beta - 2)^2}\right)$$

and

$$b_1 = \max\left\{\frac{1}{x_{\beta-1,\alpha+1}^2} + \frac{2\beta}{(\alpha + \beta)x_{\beta-1,\alpha+1}}; \frac{1}{x_{\alpha+1,\beta-1}^2} + \frac{2\alpha}{(\alpha + \beta)x_{\alpha+1,\beta-1}}\right\} \\ \leq \frac{(\alpha + \beta - 2)^2}{\min(\alpha, \beta - 2)^2} + 2\left(\frac{\alpha + \beta - 2}{\alpha + \beta}\right) \max\left(1; \frac{\beta}{\beta - 2}\right).$$

Case 4: $\alpha > 2$ and $\beta > 2$. In this case, from Lemma 3.3, $u^\alpha(1-u)^{\beta-2}$ is increasing on $[0, x_{\alpha+1,\beta-1}]$ and $u^{\alpha-2}(1-u)^\beta$ is decreasing on $[x_{\alpha-1,\beta+1}, 1]$. Noting that

$$x_{\alpha-1,\beta+1} < x_{\alpha,\beta} < x_{\alpha+1,\beta-1},$$

setting $x_* = x_{\alpha,\beta}$, by (3.15) and (3.16), we obtain

$$b_0 = \max\left(\frac{\alpha - 1}{x_{\alpha,\beta}^2}; \frac{\beta - 1}{x_{\beta,\alpha}^2}\right) = (\alpha + \beta - 2)^2 \max\left(\frac{1}{\alpha - 1}; \frac{1}{\beta - 1}\right)$$

and

$$b_1 = \max\left\{\frac{1}{x_{\alpha,\beta}^2} + \frac{2\alpha}{(\alpha + \beta)x_{\alpha,\beta}}; \frac{1}{x_{\beta,\alpha}^2} + \frac{2\beta}{(\alpha + \beta)x_{\beta,\alpha}}\right\} \\ \leq \frac{(\alpha + \beta - 2)^2}{\min(\alpha - 1, \beta - 1)^2} + 2\left(\frac{\alpha + \beta - 2}{\alpha + \beta}\right) \max\left(\frac{\alpha}{\alpha - 1}; \frac{\beta}{\beta - 1}\right).$$

For the final inequality in (3.13), with $p(y; \alpha, \beta)$ denoting the $\mathcal{B}(\alpha, \beta)$ density in (1.2), we have

$$\begin{aligned} \|h - \mathcal{B}h\| &= \sup_{x \in [0,1]} |h(x) - \mathcal{B}h| \\ &\leq \sup_{x \in [0,1]} \int_0^1 |h(x) - h(y)| p(y; \alpha, \beta) dy \\ &\leq \|h'\| \sup_{x \in [0,1]} \int_0^1 |x - y| p(y; \alpha, \beta) dy \\ &\leq \|h'\|. \end{aligned}$$

4. Distance of the distribution of L_{2n} to the arcsine law

We rely on [11] in the following proof of Theorem 1.2, noting that in [11] no explicit bound is obtained.

Proof. Let p be the mass function of L_{2n} given by (1.3), and let Z have the arcsine distribution. Applying Proposition 2.1 for p , followed by Corollary 2.1 with

$$\psi(k) = \frac{2k - n + 1}{(k + 1)(2(n - k) - 1)} \quad \text{and the choice} \quad c(k) = (k + 1)(2(n - k) - 1),$$

Döbler [11] arrived at the version of Lemma 2.1 for this case, showing that $W_n = (2n)^{-1}L_{2n}$ is so distributed if and only if

$$\mathbb{E} \left[nW_n \left(1 - W_n + \frac{1}{2n} \right) \Delta_{1/n} f \left(W_n - \frac{1}{n} \right) + \left(\frac{1}{2} - W_n \right) f(W_n) \right] = 0 \tag{4.1}$$

for all functions $f \in \mathcal{F}(p)$.

Now following the steps given in Theorem 1.1 for the Pólya urn, collecting the estimates from the proof of Theorem 3.1 of [11] shows that if f is the solution (3.2) to (2.3) with $\alpha = \beta = \frac{1}{2}$, so that $b_0 = 2$ and $b_1 = 6$, then, for all differentiable functions h , we have

$$|\mathbb{E}h(W_n) - \mathbb{E}h(Z)| \leq \frac{1}{2n} \|h'\| + \frac{1}{2n} \|f\| + \left(\frac{1}{4n} + \frac{n+2}{2n^2} + \frac{3}{4n} \right) \|f'\|.$$

Applying the bounds of Lemma 3.4 as well as Lemma 3.2 yields the bound in Theorem 1.2.

Applying (4.1) with $f(w)$ replaced by $g(w) = 1$ and $g(w) = w$ yields

$$\mathbb{E}W_n = \frac{1}{2} \quad \text{and} \quad \mathbb{E}W_n^2 = \frac{3}{8} + \frac{1}{8n},$$

respectively. For $f(w) = w^2/2$, a function in $\text{Lip}(1)$ on $[0, 1]$, since $\mathbb{E}Z^2 = \frac{3}{8}$, we obtain

$$\mathbb{E}f(W_n) - \mathbb{E}f(Z) = \frac{1}{16n}.$$

Hence, the $O(1/n)$ rate cannot be improved.

Acknowledgements

We would like to thank the Keble Advanced Studies Centre, Oxford, for support and an anonymous referee for helpful comments which led to an improvement of the paper. Larry Goldstein acknowledges the support of the NSA under grant H98230-11-1-0162.

References

- [1] ARGIENTO, R., PEMANTLE, R., SKYRMS, B. AND VOLKOV, S. (2009). Learning to signal: analysis of a micro-level reinforcement model. *Stoch. Process. Appl.* **119**, 373–390.
- [2] ARRATIA, R., GOLDSTEIN, L. AND GORDON, L. (1989). Two moments suffice for Poisson approximations: the Chen-Stein method. *Ann. Prob.* **17**, 9–25.
- [3] BARBOUR, A. D. (1990). Stein’s method for diffusion approximations. *Prob. Theory Relat. Fields* **84**, 297–322.
- [4] BARBOUR, A. D., HOLST, L. AND JANSON, S. (1992). *Poisson Approximation*. Oxford University Press, New York.
- [5] BROWN, T. C. AND PHILLIPS, M. J. (1999). Negative binomial approximation with Stein’s method. *Methodology Comput. Appl. Prob.* **1**, 407–421.
- [6] CHATTERJEE, S., FULMAN, J. AND RÖLLIN, A. (2011). Exponential approximation by Stein’s method and spectral graph theory. *ALEA Lat. Amer. J. Prob. Math. Statist.* **8**, 197–223.
- [7] CHAUVIN, B., POUYANNE, N. AND SAHNOUN, R. (2011). Limit distributions for large Pólya urns. *Ann. Appl. Prob.* **21**, 1–32.
- [8] CHEN, L. H.-Y., GOLDSTEIN, L. AND SHAO, Q.-M. (2011). *Normal Approximation by Stein’s Method*. Springer, Heidelberg.
- [9] CHU, W. (2007). Abel’s lemma on summation by parts and basic hypergeometric series. *Adv. Appl. Math.* **39**, 490–514.
- [10] CHUNG, F., HANDJANI, S. AND JUNGREIS, D. (2003). Generalizations of Pólya’s urn problem. *Ann. Combinatorics* **7**, 141–153.
- [11] DÖBLER, C. (2012). A rate of convergence for the arcsine law by Stein’s method. Preprint. Available at <http://arxiv.org/abs/1207.2401v1>.
- [12] DÖBLER, C. (2012). Stein’s method of exchangeable pairs for absolutely continuous, univariate distributions with applications to the Polya urn model. Preprint. Available at <http://arxiv.org/abs/1207.0533v2>.
- [13] DURRETT, R. (2007). *Random Graph Dynamics*. Cambridge University Press.
- [14] EICHELSBACHER, P. AND REINERT, G. (2008). Stein’s method for discrete Gibbs measures. *Ann. Appl. Prob.* **18**, 1588–1618.
- [15] FELLER, W. (1970). *An Introduction to Probability Theory and Its Applications*. Vol. I. John Wiley, New York.
- [16] FISHER, R. A. (1930). *The Genetical Theory of Natural Selection*. Clarendon Press, Oxford.
- [17] GIBBS, A. L. AND SU, F. E. (2002). On choosing and bounding probability metrics. *Internat. Statist. Rev.* **70**, 419–435.
- [18] HOLMES, S. (2004). Stein’s method for birth and death chains. In *Stein’s Method: Expository Lectures and Applications*, eds P. Diaconis and S. Holmes, Institute of Mathematical Statistics, Beachwood, OH, pp. 45–67.
- [19] LEY, C. AND SWAN, Y. (2013). Local Pinsker inequalities via Stein’s discrete density approach. *IEEE Trans. Inf. Theory* **59**, 5584–5591.
- [20] LOH, W. L. (1992). Stein’s method and multinomial approximation. *Ann. Appl. Prob.* **2**, 536–554.
- [21] LUK, H.-M. (1994). Stein’s method for the gamma distribution and related statistical applications. Doctoral Thesis, University of Southern California.
- [22] MAHMOUD, H. M. (2003). Pólya urn models and connections to random trees: a review. *J. Iranian Statist. Soc.* **2**, 53–114.
- [23] MAHMOUD, H. M. (2008). *Pólya Urn Models*. Chapman & Hall/CRC Press, Boca Raton, FL.
- [24] NOURDIN, I. AND PECCATI, G. (2011). Stein’s method on Wiener chaos. *Prob. Theory Relat. Fields* **145**, 75–118.
- [25] OLVER, F. W. J., LOZIER, D. W., BOISVERT, R. F. AND CLARK, C. W. (eds) (2010). *NIST Handbook of Mathematical Functions*. Cambridge University Press.
- [26] PEKÖZ, E. (1996). Stein’s method for geometric approximation. *J. Appl. Prob.* **33**, 707–713.
- [27] PEKÖZ, E. AND RÖLLIN, A. (2011). Exponential approximation for the nearly critical Galton–Watson process and occupation times of Markov chains. *Electron. J. Prob.* **16**, 1381–1393.
- [28] PEKÖZ, E. A. AND RÖLLIN, A. (2011). New rates for exponential approximation and the theorems of Rényi and Yaglom. *Ann. Prob.* **39**, 587–608.
- [29] PEKÖZ, E. A., RÖLLIN, A. AND ROSS, N. (2013). Degree asymptotics with rates for preferential attachment random graphs. *Ann. Appl. Prob.* **23**, 1188–1218.
- [30] PEKÖZ, E. A., RÖLLIN, A. AND ROSS, N. (2013). Generalized gamma approximation with rates for urns, walks and trees. Preprint. Available at <http://arxiv.org/abs/1309.4183>.
- [31] PEMANTLE, R. (2007). A survey of random processes with reinforcement. *Prob. Surveys* **4**, 1–79.
- [32] REINERT, G. (2005). Three general approaches to Stein’s method. In *An Introduction to Stein’s Method* (Lecture Notes Ser. Inst. Math. Sci. Nat. Univ. Singapore **4**), Singapore University Press, pp. 183–221.
- [33] STEIN, C. (1972). A bound for the error in the normal approximation to the distribution of a sum of dependent random variables. In *Proc. 6th Berkeley Symp. Math. Statist. Prob.*, Vol II, University of California Press, Berkeley, CA, pp. 586–602.

- [34] STEIN, C., DIACONIS, P., HOLMES, S. AND REINERT, G. (2004). Use of exchangeable pairs in the analysis of simulations. In *Stein's Method: Expository Lectures and Applications*, eds P. Diaconis and S. Holmes, Institute of Mathematical Statistics, Beachwood, OH, pp. 1–26.
- [35] WILCOX, R. R. (1981). A review of the beta-binomial model and its extensions. *J. Educational Behavioral Statist.* **6**, 3–32.
- [36] WRIGHT, S. (1945). The differential equation of the distribution of gene frequencies. *Proc. Nat. Acad. Sci. USA* **31**, 382–389.
- [37] WRIGHT, S. (1949). Adaptation and selection. In *Genetics, Paleontology and Evolution*, eds G. Jepson, G. Simpson and E. Mayr, Princeton University Press, pp. 365–389.