Unbalanced multi-drawing urn with random addition matrix

Aguech Rafik
Department of Statistics and Operation Research, King Saud University, Riyadh, Saudi Arabia, and
Selmi Olfa
University of Monastir, Monastir, Tunisia

Received 11 October 2018
Revised 23 December 2018
Accepted 27 December 2018

Abstract
In this paper, we consider a two color multi-drawing urn model. At each discrete time step, we draw uniformly at random a sample of m balls ($m \geq 1$) and note their color, they will be returned to the urn together with a random number of balls depending on the sample's composition. The replacement rule is a 2×2 matrix depending on bounded discrete positive random variables. Using a stochastic approximation algorithm and martingales methods, we investigate the asymptotic behavior of the urn after many draws.

Keywords Central limit theorem, Unbalanced urn, Martingale, Stochastic algorithm
Paper type Original Article

1. Introduction

The classical Pólya urn was introduced by Pólya and Eggenberger [7] describing contagious diseases. The first model is as follows: An urn contains balls of two colors at the start, white and blue. At each step, one picks a ball randomly and returns it to the urn with a ball of the same color. Afterwards, there were many generalizations and urn model become a simple tool to describe several models such finance, clinical trials (see [19,22]), biology (see [11]), computer sciences, internet (see [8,18]), etc...

Recently, Mahmoud, Chen, Wei, Kuba and Sulzbach [4,5,12–15], have focused on the multi-drawing urn. Instead of picking a ball,one picks a sample of m balls ($m \geq \ell$), say ℓ white and $(m-\ell)$ blue balls. The pick is returned back to the urn together with $a_{m-\ell}$ white and $b_{m-\ell}$ blue balls, where a_ℓ and b_ℓ, $0 \leq \ell \leq m$ are integers. At first, they treated two particular cases when $\{a_{m-\ell} = c \times \ell$ and $b_{m-\ell} = c \times (m-\ell)\}$ and when $\{a_{m-\ell} = c \times (m-\ell)$ and $b_{m-\ell} = c \times \ell\}$, where c is a positive constant. By different methods as martingales and moment methods, the

The first author is grateful to the King Saud University, Deanship of Scientific Research, College of Science Research Center. The authors also thank two anonymous referees for their valuable comments and suggestions.

The publisher wishes to inform readers that the article "Unbalanced multi-drawing urn with random addition matrix" was originally published by the previous publisher of the *Arab Journal of Mathematical Sciences* and the pagination of this article has been subsequently changed. There has been no change to the content of the article. This change was necessary for the journal to transition from the previous publisher to the new one. The publisher sincerely apologises for any inconvenience caused. To access and cite this article, please use Rafik, A., Olfa, S. (2019), "Unbalanced multi-drawing urn with random addition matrix" *Arab Journal of Mathematical Sciences*, Vol. 26 No. 1/2, pp. 57-74. The original publication date for this paper was 11/01/2019.

Arab Journal of Mathematical Sciences
Vol. 26 No. 1/2, 2020
pp. 57-74
Emerald Publishing Limited
e-ISSN: 2588-9214
p-ISSN: 1319-5166
DOI 10.1016/j.ajmsc.2018.12.004

authors described the asymptotic behavior of the urn composition. When considering the general case and in order to ensure the existence of a martingale, they supposed that W_n, the number of white balls in the urn after n draws, satisfies the affinity condition i.e, there exist two deterministic sequences (α_n) and (β_n) such that, for all $n \geq 0$, $\mathbb{E}[W_{n+1}|\mathcal{F}_n] = \alpha_n W_n + \beta_n$. Under this condition, the authors focused on small and large index urns. Later, the affinity condition was removed in the work of Lasmer, Mailler and Selmi [16], they generalized this model and looked at the case of more than two colors.

This paper contains the first results about multi drawing Pólya urns with random replacement rule. Even in the classical Pólya urn, where one ball is picked at every time step very few results cover the unbalanced case: exceptions are the works of Janson and Aguech. In [9] Janson studied a generalized urn model containing q different colors ($q \geq 1$) with a $q \times q$ replacement matrix A with random entries such that $A_{i,j} \geq 0$ and $\mathbb{E}(A_{i,j}^2) < \infty$ for all $i, j = 1, \ldots, q$. Janson considered the case when the mean of A is an irreducible matrix. Using the method of embedding in continuous time of Athrea and Karlin [3], he gave explicit formulas for the asymptotic variances and covariances as well as functional limit theorems for the urn. Then, Janson [10] considered a particular two color Pólya urn model evolving according to a triangular replacement matrix (the matrix in non irreducible) with deterministic entries. He established theorems describing the asymptotic behavior of the composition of the urn after n draws. Afterwards, Aguech [1] extended some results and studied two colors urn model with triangular replacement matrix. The entries of such a matrix, X_n, Y_n and C_n, are positive random variables with finite means and variances. The embedding in continuous times' method were successful once again and he gave theorems about the asymptotic behavior of the urn's composition after a long time.

In this paper, we deal with a two color unbalanced urn class with multiple drawing and random addition matrix. Consider X and Y two discrete-valued random variables. We assume that there exists two constants $U > 0$ and $L > 0$ such that $L \leq X \leq U$ and $L \leq Y \leq L$. Let $(X_n)_{n\geq 0}$ (resp $(Y_{n\geq 0})_{n\geq 0}$) be a sequence of independent random variables distributed like X (resp Y). The sequences X_n and Y_n are not assumed to be independent.

The model we study is defined as follows: An urn contains initially W_0 white balls and B_0 blue balls, we fix an integer $m \geq 1$, at a discrete step $n \geq 1$, we draw uniformly at random a sample of m balls, we denote by ξ_n the number of white balls among those m balls (we assume that the initial composition of the urn is more than m to make the first draw possible). We return the drawn sample together with $Q_n(\xi_n, m - \xi_n)^t$ balls, where Q_n is a 2×2 matrix depending on the random variables X_n and Y_n. Let us denote by W_n (resp B_n) the number of white balls (resp blue balls), T_n the total number of balls and by Z_n the proportion of white balls in the urn at time n. In other words, the process is defined recursively as follows: for all $n \geq 1$

$$\begin{pmatrix} W_n \\ B_n \end{pmatrix} \stackrel{\mathcal{D}}{=} \begin{pmatrix} W_{n-1} \\ B_{n-1} \end{pmatrix} + Q_n \begin{pmatrix} \xi_n \\ m - \xi_n \end{pmatrix}. \tag{1}$$

Let $\mathcal{F}_n$ be the σ-field generated by the first n draws. Note that, with these notations, we have for $k \in \{0, \ldots, m\}$,

$$\mathbb{P}[\xi_n = k|\mathcal{F}_{n-1}] = \frac{\binom{W_{n-1}}{k}\binom{B_{n-1}}{m-k}}{\binom{T_{n-1}}{m}}. \tag{2}$$

Thus, conditioning on $\mathcal{F}_{n-1}$ the variable ξ_n has an hypergeometric distribution with parameters m, Z_{n-1} and T_{n-1}. Some particular cases were the interest of recent works [4,15] and [2], where the authors characterized the urn models defined by Eq. (1) for the following cases

$$Q_n \in \left\{ \begin{pmatrix} a & 0 \\ 0 & a \end{pmatrix}, \begin{pmatrix} 0 & a \\ a & 0 \end{pmatrix}, \begin{pmatrix} a & 0 \\ 0 & b \end{pmatrix}, \begin{pmatrix} 0 & a \\ b & 0 \end{pmatrix} \right\},$$

where a, b are strictly positive integers. To generalize the previous works, we consider the urn models evolving according to Eq. (1) with

$$Q_n \in \left\{ \begin{pmatrix} X_n & 0 \\ 0 & Y_n \end{pmatrix}, \begin{pmatrix} 0 & X_n \\ Y_n & 0 \end{pmatrix}, \begin{pmatrix} X_n & 0 \\ 0 & X_n \end{pmatrix}, \begin{pmatrix} 0 & X_n \\ X_n & 0 \end{pmatrix} \right\}.$$

The main idea is to use the stochastic algorithms and martingales in order to prove that the number of white balls in the urn converges almost surely and to study its fluctuations around its limit whenever it is possible.

The paper is organized as follows. In Section 2, we give the main results of the paper. Section 3 is devoted to the details of the stochastic approximation algorithm's method. The proofs of the main results are detailed in Section 4.

2. Main results

We start with some notations. The notation $a.s.$ stands for *almost surely*. For a random variable R, we denote by

$$\mu_R = \mathbb{E}(R) \text{ and } \sigma_R^2 = \mathbb{V}ar(R),$$

by $\mu_X := \mu_{X_1}$ (respectively $\mu_Y := \mu_{Y_1}$) and $\sigma_X^2 := \sigma_{X_1}^2$ (respectively $\sigma_Y^2 := \sigma_{Y_1}^2$). For x_n and y_n two sequences of real numbers such that $y_n \neq 0$ for all n, we denote $x_n = o(y_n)$ (respectively $x_n = o(y_n)$, $a.s$) if $\lim_{n\to+\infty} x_n/y_n = 0$ (if $\lim_{n\to+\infty} x_n/y_n = 0$, $a.s$ when x_n and y_n are random).

In this section we state our main result. As mentioned in the introduction, we study urn models evolving according to Eq. (1). Recall that in the whole of paper we consider $(X_n)_{n\geq1}$ (resp $(Y_n)_{n\geq1}$), a sequence of independent random variables distributed like X (resp Y).

The present theorem deals with an urn evolving with an anti-diagonal replacement matrix. The model is then opposite reinforced, i.e the more color is drawn the more it reinforces the opposite color.

Theorem 1. *Let* $z := \frac{\sqrt{\mu_X}}{\sqrt{\mu_X}+\sqrt{\mu_Y}}$ *and consider the urn model evolving by the matrix* $Q_n = \begin{pmatrix} 0 & X_n \\ Y_n & 0 \end{pmatrix}$. *We have the following results*:

(1) The total number of balls in the urn after n draws satisfies

$$T_n = \sqrt{\mu_X \mu_Y}\, m\, n + o(n),\ a.s. \tag{3}$$

and the number of white and blue balls in the urn after n draws satisfy

$$W_n = \mu_X (1-z)\, m\, n + o(n),\ a.s.$$

$$B_n = \mu_Y z\ m\ n + o(n),\ a.s.$$

(2) Furthermore, with $G(x) = \sum_{i=0}^4 a_i x^i$, *the normalized number of white balls in the urn satisfies the central limit theorem*

$$\frac{W_n - zT_n}{\sqrt{n}} \xrightarrow{\mathcal{D}} \mathcal{N}\left(0, \frac{G(z)}{3}\right), \quad as\ \ n \to +\infty. \tag{4}$$

(3) *Furthermore, when* $Y_n = X_n$ *for all* $n \geq 0$*, the total number of balls in the urn after n draws satisfies, for any* $\delta > \frac{1}{2}$

$$T_n = m\mu_X n + o(\sqrt{n}\ \ln^{\delta} n), \quad a.s.$$

The number of white balls W_n *and blue balls* B_n *in the urn after n draws satisfy for any* $\delta > \frac{1}{2}$,

$$W_n = \frac{m\mu_X}{2} n + o(\sqrt{n}\ \ln^{\delta} n), \quad a.s,$$

$$B_n = \frac{m\mu_X}{2} n + o(\sqrt{n}\ \ln^{\delta} n), \quad a.s.$$

We have the convergence in distribution:

$$\lim_{n\to+\infty} \frac{W_n - \frac{1}{2}T_n}{\Sigma\sqrt{n}} = \mathcal{N}(0,\ 1) \ \ and \ \ \lim_{n\to+\infty} \frac{W_n - \mathbb{E}(W_n)}{\Sigma_1\sqrt{n}} = \mathcal{N}(0,\ 1);$$

where

$$\Sigma = \frac{m}{12}\left(\sigma_X^2 + \mu_{X^2}\right) \ \ and \ \ \Sigma_1 = \frac{m}{12}\left[\left(\sigma_X^2 + \mu_{X^2}\right) + m^2\sigma_X^2\right].$$

Example 1. Let $X_n = a$ and $Y_n = b$ (where a and b are not random), then $z = \frac{\sqrt{a}}{\sqrt{a}+\sqrt{b}}$. This case was studied in [2] and the authors proved the following

$$\sqrt{n}\left(\frac{W_n}{T_n} - z\right) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \frac{\sqrt{ab}}{3m(\sqrt{a} + \sqrt{b})^2}), \quad \text{as } n \to \infty.$$

Under the notation of Theorem 1, we easily compute $G(z) = mabz(1-z)$ and then the particular case is proved again.

Example 2. Let $X_n = Y_n = C$ (non random), the urn is balanced and the total number of balls is deterministic and satisfies $T_n = T_0 + Cmn$. Furthermore, we have $\mu_X = C$ and $\sigma_X^2 = 0$, applying Theorem 1(3) we obtain the following limit:

$$\frac{W_n - \frac{Cmn}{2}}{\sqrt{n}} \xrightarrow{\mathcal{D}} \mathcal{N}(0, \frac{mC^2}{12}), \quad \text{as } n \to \infty.$$

Kuba et al. [15] studied this particular case and established such a result via two different methods: The recursion formulas permit to derive the expression of the higher moments of the number of white balls and then to conclude functional limit theorem. The same result was proved via martingales method.

In the following theorem, we consider a diagonal replacement matrix Q_n. The model is self reinforced since the rich gets richer. As the particular case when $m = 1$, we compare $\frac{\mu_X}{\mu_Y}$ with 1, we will distinguish different phases.

Theorem 2. *Consider the urn evolving by the matrix* $Q_n = \begin{pmatrix} X_n & 0 \\ 0 & Y_n \end{pmatrix}$.

(1) *If* $\mu_X > \mu_Y$*, then the total number of balls in the urn after n draws satisfies*

$$T_n = m\mu_X n + o(n), \quad a.s.,$$

and the asymptotic composition of the urn is

$$W_n = m\mu_X n + o(n), \quad B_n = B_\infty n^\rho + o(n^\rho), \quad a.s.$$

where $\rho = \frac{\mu_Y}{\mu_X}$ *and* B_∞ *is a positive random variable.*

(2) *If* $\mu_X = \mu_Y$, *the composition of the urn after n draws satisfies*

$$T_n = m\mu_X n + o(n), \quad a.s.$$

In addition, there exists a positive random variable W_∞ *such that,*

$$W_n = W_\infty n + o(n) \text{ and } B_n = (\mu_x m - W_\infty)\, n + o(n), \quad a.s.$$

(3) *Furthermore, if for all* $n \geq 0$, $Y_n = X_n$, *the distribution of the random variable* W_∞ *is absolutely continuous.*

Remark. The case when $\mu_X < \mu_Y$ is obtained by interchanging the colors. In fact we have the following almost sure results:

$$T_n = m\mu_Y n + o(n),\ W_n = W_\infty n^\sigma + o(n) \text{ and } B_n = m\mu_Y n + o(n),$$

where W_∞ is a positive random variable and $\sigma = \frac{\mu_X}{\mu_Y}$.

Example 3. Aguech [1] studied the particular case when $m = 1$ and considered the following triangular replacement matrix

$$\begin{pmatrix} X_n & 0 \\ C_n & Y_n \end{pmatrix},$$

where X_n, Y_n and C_n are independent positive random variables with finite means and variances. Via embedding in continuous time method and martingales, the author proved, for $C_n = 0$, the following almost sure results:

(a) If $\mu_X > \mu_Y$,

$$W_n = \mu_X n + o(n), B_n = Dn^\rho \text{ and } T_n = \mu_X n + o(n),$$

where $\rho = \frac{\mu_Y}{\mu_X}$ and D is a positive random variable.

(b) If $\mu_X = \mu_Y$,

$$W_n = \mu_X \frac{W}{W+B} n + o(n) \text{ and } B_n = \mu_X \frac{B}{W+B} n + o(n),$$

where W and B are the almost sure limit of a continuous time martingale.

We prove again these results in Theorem 2 using stochastic approximation algorithm.

Example 4. Chen and Kuba [4] studied the case when $X_n = Y_n = C$ (C is non random) and $m \geq 1$. They gave explicit expressions of moment of all order of W_n/n and proved that its almost sure limit, W_∞ cannot be an ordinary *Beta* distribution, unlike the original Pòlya urn model [7] when $X = C$ and $m = 1$, Eggenberger and Pólya proved in 1923 that the random variable W_∞/C has a Beta distribution with parameters $(B_0/C,\ W_0/C)$. Unfortunately, in our model we cannot yet derive the expression of higher moments of W_n/n since the recurrence formulas are too intricate.

3. Some results on stochastic approximation algorithm

The stochastic algorithm approximation plays a crucial role in the proofs in order to describe the asymptotic composition of the urn. As many versions of the stochastic algorithm exist in the literature (see [6] for example), we adapt the version of Renlund in [20,21].

Definition 1. A stochastic approximation algorithm $(U_n)_{n\geq 0}$ is a stochastic process taking values in $[0,1]$ and adapted to a filtration $\mathcal{F}_n$ that satisfies

$$U_{n+1} - U_n = \gamma_{n+1}(f(U_n) + \Delta M_{n+1}), \tag{5}$$

where $(\gamma_n)_{n\geq 1}$ and $(\Delta M_n)_{n\geq 1}$ are two $\mathcal{F}_n$-measurable sequences of random variables, f is a function from $[0, 1]$ into $\mathbb{R}$ such that $f(0)\geq 0$, $f(1)\leq 0$ and the following conditions hold almost surely: There exists constants c_1, c_2, K_Δ, and K_f positive real numbers such that for any $n\geq 1$,

(i) $\frac{c_1}{n} \leq \gamma_n \leq \frac{c_2}{n}$;

(ii) $\mathbb{E}((\Delta M_{n+1})^2|\mathcal{F}_n) \leq K_\Delta$;

(iii) $|f(U_n)| \leq K_f$;

(iv) $\mathbb{E}[\gamma_{n+1}\Delta M_{n+1}|\mathcal{F}_n] = 0$.

Definition 2. Let $\mathcal{Z}_f = \{x\in[0, 1]; f(x) = 0\}$. A zero $p\in\mathcal{Z}_f$ will be called stable if there exists a neighborhood $\mathcal{N}_p$ of p such that $f(x)(x-p) < 0$ whenever $x\in\mathcal{N}_p\backslash\{p\}$. If f is differentiable, then $f'(p)$ is sufficient to determine that p is stable.

Remark. Note that Assumption (ii) in Definition 1 is not stated as in [20] where it is assumed that there exists a positive constant K_Δ such that $|\Delta M_n|\leq K_\Delta$.

We have the following result about the process defined by Eq. (5)

Proposition 1. *Let $(U_n)_{n\geq 0}$ be a stochastic algorithm defined by* Eq. (5). *If f is continuous, then* $\lim_{n\to+\infty} U_n$ *exists almost surely and is a stable zero of f.*

The following lemmas will be useful for the proof of Proposition 1.

Lemma 1. *Define $V_n = \sum_{i=1}^n \gamma_i\Delta M_i$. Under the assumptions of Proposition 1, V_n converges almost surely.*

Proof. Under the assumptions mentioned in Definition 1, we have

$$\mathbb{E}(V_{n+1}|\mathcal{F}_n) = V_n + \mathbb{E}(\gamma_{n+1}\Delta M_{n+1}|\mathcal{F}_n) = V_n.$$

We deduce that $(V_n, \mathcal{F}_n)_n$ is a martingale. On the other hand,

$$\mathbb{E}\left(V_n^2\right) = \sum_{i=1}^n \mathbb{E}\left(\gamma_i^2(\Delta M_i)^2\right) \leq \sum_{i=1}^n \frac{c_2^2}{i^2}\mathbb{E}\left((\Delta M_i)^2\right) \leq K_\Delta c_2^2 \sum_{i=1}^n \frac{1}{i^2} < \infty.$$

It follows that $(V_n)_n$ is an $\mathbb{L}^2$- bounded martingale, and thus, it converges almost surely. □

Next lemma ensures that, under the assumptions of Proposition 1, all possible candidates for the almost sure limit of U_n are necessary among the zeros of f.

Lemma 2 ([20]). *Let $\mathcal{Z}_f = \{x\,;\ f(x) = 0\}$ be the set of zeros of f and let $\mathcal{C}(U_n)$ be the set of limit points of $\{U_n\}$ defined by*

$$\mathcal{C}(U_n) = \bigcap_{n\geq 1}\overline{\{U_n, U_{n+1}, \ldots\}},$$

where $\overline{A}$ denotes the closure of a set A. Under the assumptions of Proposition 1, if f is continuous, then,

$$\mathbb{P}(\mathcal{C}(U_n) \subseteq \mathcal{Z}_f) = 1.$$

Lemma 3 (*[20]*). *Suppose that* $f(x) < -\delta$ *(or* $f(x) > \delta$*) for some* $\delta > 0$*, whenever* $x \in (a_0, \; b_0)$*. Then,*

$$\mathcal{C}(U_n)\bigcap(a_0, b_0) = \emptyset \quad a.s.,$$

and either $\limsup_n U_n \leq a_0$ *or* $\liminf_n U_n \geq b_0$.

We are now able to handle the proof of Proposition 1.

Proof of Proposition 1. The proof is close to Theorem 1 in [20], for the convenience of the reader, we resume the proof and we mention the main steps. If $\lim_{n\to+\infty} U_n$ does not exist, we can find two rational numbers in the open interval

$]\liminf_{n\to+\infty} U_n, \; \limsup_{n\to+\infty} U_n[$. Let $\liminf U_n < p < q < \limsup U_n$ be two arbitrary different rational numbers. If we can show that

$$\mathbb{P}\big(\{\liminf U_n \leq p\}\cap\{\limsup U_n \geq q\}\big) = 0,$$

then, the existence of the limit will be established and the claim of the proposition follows from Lemma 2. For this reason, we need to distinguish two different cases whether or not p and q are in the same connected component of $\mathcal{Z}_f$.

Case 1: p **and** q **are not in the same connected component of** $\mathcal{Z}_f$: Since $\mathcal{Z}_f$ is closed and f is continuous there must exist $[a, b] \subseteq [p, q]\bigcap\mathcal{Z}_f^c$ such that f is non-zero and has a constant sign for all $x \in (a, \; b)$. By Lemma 3, it is impossible to have $\liminf_n U_n \leq a$ and $\limsup_n U_n \geq b$.

Case 2: p **and** q **are in the same connected component of** $\mathcal{Z}_f$: In all the cases of our framework $\mathcal{Z}_f$ is a set of two isolated points, therefore we are not interested to the case when p and q are not in the same connected component.

To establish that the almost sure limit of U_n is among the stable point set, we refer the reader to [20] to see a detailed proof. □

Next result is due to Renlund [21] which will be used in the proofs of Theorems 1 and 2.

Theorem 3 (*[21]*). *Let* $(U_n)_{n\geq 0}$ *satisfy Eq. (5) and that* $\lim_{n\to+\infty} U_n = U^*$*. Let*

$$\widehat{\gamma}_n := n\gamma_n \widehat{f}(U_{n-1}), \quad where \; \widehat{f}(x) = \frac{f(x)}{U^* - x}.$$

If $\widehat{\gamma}_n$ *converges almost surely to some limit* $\widehat{\gamma} > \frac{1}{2}$ *and if* $E[(n\gamma_n \Delta M_n)^2 | \mathcal{F}_{n-1}] \to \sigma^2 > 0$*, then, we have the convergence in distribution*

$$\sqrt{n}(U_n - U^*) \xrightarrow{\mathcal{D}} \mathcal{N}\left(0, \; \frac{\sigma^2}{2\widehat{\gamma} - 1}\right).$$

4. Proof of the main results

4.1 Prerequisite for the proofs of the main results

We show in the following that the stochastic approximation algorithm is a fruitful method to study unbalanced urn models. Although there are few versions of such a method that permit to γ_n to be random, the version of Renlund [20] and [21] applies to our model.

Under the assumptions of Theorem 1 and according to Eq. (1), the compositions of the urn satisfy the following recursions:

$$W_{n+1} = W_n + X_{n+1}(m - \xi_{n+1}) \tag{6}$$

and

$$T_{n+1} = T_n + mX_{n+1} + \xi_{n+1}(Y_{n+1} - X_{n+1}). \tag{7}$$

We start with first results that will be useful for the proof of Theorem 2.

Lemma 4 (*Technical Lemma*). *For all integers m, A, B such that $m \leq A + B$ we have*

$$\sum_{m=0}^{m} k \binom{A}{k} \binom{B}{m-k} = A \binom{A+B-1}{m-1}$$

and

$$\sum_{m=0}^{m} k^2 \binom{A}{k} \binom{B}{m-k} = A(A-1) \binom{A+B-2}{m-2} + A \binom{A+B-1}{m-1}$$

Remark. Since conditioning on $\mathcal{F}_{n-1}$ the variable (ξ_n) has an hypergeometric distribution with parameters m, Z_{n-1} and T_{n-1}, it follows from Lemma 4 the following:

$$\mathbb{E}(\xi_n | \mathcal{F}_{n-1}) = mZ_n,$$

and

$$\mathbb{V}ar(\xi_n | \mathcal{F}_{n-1}) = mZ_{n-1}(1 - Z_{n-1}) \frac{T_{n-1} - m}{T_{n-1} - 1}.$$

Lemma 5. *Under the assumptions of Theorem 1, the proportion of white balls after n draws, Z_n, satisfies the stochastic algorithm defined by (5), where $\gamma_n = \frac{1}{T_n}$,*

$$f(x) = m(\mu_X - \mu_Y)x^2 - 2\mu_X mx + \mu_X m,$$

and

$$\Delta M_{n+1} = D_{n+1} - \mathbb{E}[D_{n+1} | \mathcal{F}_n],$$

with

$$D_{n+1} = \xi_{n+1}(Z_n(X_{n+1} - Y_{n+1}) - X_{n+1}) + mX_{n+1}(1 - Z_n).$$

Proof. In view of the recursions in Equations (6), (7) we have

$$\begin{aligned} Z_{n+1} - Z_n &= \frac{1}{T_{n+1}} \left[W_n + X_{n+1}(m - \xi_{n+1}) - Z_n(T_n + mX_{n+1} + \xi_{n+1}(Y_{n+1} - X_{n+1})) \right] \\ &= \frac{1}{T_{n+1}} \left[X_{n+1}(m - \xi_{n+1}) - Z_n(mX_{n+1} + \xi_{n+1}(Y_{n+1} - X_{n+1})) \right] \\ &= \frac{D_{n+1}}{T_{n+1}}. \end{aligned}$$

An easy computation shows that $\mathbb{E}(D_{n+1} | \mathcal{F}_n) = m(\mu_X - \mu_Y)Z_n^2 - 2m\mu_X Z_n + m\mu_X$. □

Using Proposition 1, we show that the almost sure limit of the proportion of white balls in the urn depends on the means of the variables X_n and Y_n:

Proposition 2. *The proportion of white balls in the urn after n draws, under the assumptions of Theorem 1, satisfies*

$$\lim_{n\to+\infty} Z_n = z := \frac{\sqrt{\mu_X}}{\sqrt{\mu_X}+\sqrt{\mu_Y}},\ a.s. \tag{8}$$

Proof. In view of Lemma 5, we check the assumptions of Definition 1, indeed,

(i) an easy computation shows that

$$T_n = T_0 + m\sum_{i=1}^{n}(m-\xi_i)X_i + \sum_{i=1}^{n}\xi_i Y_i. \tag{9}$$

Since for all $n \geq 1$ we have $0 \leq \xi_n \leq m, L \leq X_n \leq U$ and $L \leq Y_n \leq U$, then

$$mnL \leq T_n \leq T_0 + mnU.$$

Then the following bound holds, for all $n \geq 1$

$$\frac{c_1}{n} \leq \frac{1}{T_n} \leq \frac{c_2}{n}, \tag{10}$$

with $c_1 = \frac{1}{T_0+mU}$ and $c_2 = \frac{1}{mL}$.

(ii)

$$\mathbb{E}\left[(\Delta M_{n+1})^2|\mathcal{F}_n\right] \leq \left(\mu_{(X-Y)^2}+3\mu_X\right)\left(m+m^2\right)+5m^2\mu_{X^2}+2m^2\mu_X\mu_Y$$
$$+m^2(|\mu_X-\mu_Y|+3\mu_X) = K_\Delta,$$

(iii) $|f(Z_n)| \leq m(|\mu_Y-\mu_X|+3\mu_X) = K_f,$

(iv) $\mathbb{E}\left[\frac{1}{T_{n+1}}\Delta M_{n+1}|\mathcal{F}_n\right] \leq \frac{1}{T_n}\mathbb{E}[\Delta M_{n+1}|\mathcal{F}_n] = 0.$

Since the function f, defined in Lemma 5, is continuous, we conclude by Proposition 1, that the process Z_n converges $a.s.$ to

$$z = \frac{\sqrt{\mu_X}}{\sqrt{\mu_X}+\sqrt{\mu_Y}},$$

which is the unique zero of f with negative derivative. □

The following Lemma will intervene in the proof of Theorem.

Lemma 6. *Under the assumptions of Theorem 1, the total number of balls after n draws satisfies*

$$\lim_{n\to+\infty}\frac{T_n}{n} = m\sqrt{\mu_X\mu_Y},\ \ a.s.$$

Proof. Let $G_n = \sum_{i=1}^{n}\left[\xi_i(Y_i-X_i) - \mathbb{E}[\xi_i(Y_i-X_i)|\mathcal{F}_{i-1}]\right]$, by the recursive Eq. (7), we have

$$\frac{T_n}{n} = \frac{T_0}{n} + \frac{m}{n}\sum_{i=1}^{n}X_i + \frac{m(\mu_Y-\mu_X)}{n}\sum_{i=1}^{n}Z_{i-1} + \frac{G_n}{n}.$$

Since $(X_i)_{i\geq 1}$ are i.i.d. random variables, then by the strong law of large numbers we have

$$\frac{m}{n}\sum_{i=1}^{n} X_i \xrightarrow{a.s} m\mu_X.$$

Via Proposition 2 and Cesáro lemma, we conclude that $\frac{1}{n}\sum_{i=1}^{n} Z_{i-1}$ converges $a.s.$, as n goes to infinity, to z. Finally, we prove that the last term in the right side tends $a.s.$ to zero, as n tends to infinity. In fact, $(G_n, \mathcal{F}_n)$ is a martingale difference sequence with quadratic variation given by

$$\langle G\rangle_n = \sum_{i=1}^{n} \mathbb{E}[(\nabla G_i)^2|\mathcal{F}_{i-1}],$$

where $\nabla G_n = G_n - G_{n-1} = \xi_n(Y_n - X_n) - \mathbb{E}[\xi_n(Y_n - X_n)|\mathcal{F}_{n-1}]$. By a simple computation, we have the almost sure convergence

$$\lim_{n\to+\infty} \mathbb{E}[(\nabla G_n)^2|\mathcal{F}_{n-1}] = \left(mz(1-z) + m^2z^2\right)\left(\sigma_Y^2 + \sigma_X^2\right).$$

Therefore, Cesáro lemma ensures that $a.s.$

$$\lim_{n\to+\infty} \frac{\langle G\rangle_n}{n} = \left(mz(1-z) + m^2z^2\right)\left(\sigma_Y^2 + \sigma_X^2\right).$$

It follows that $\frac{G_n}{n} \xrightarrow{a.s} 0$. Thus, for n large enough, we have

$$\frac{T_n}{n} \xrightarrow{a.s} m\sqrt{\mu_X\mu_Y}. \quad \square \tag{11}$$

Remark. The convergence in Proposition 2 holds also in $\mathbb{L}^2$.

Under the hypothesis of Theorem 2, the process of the urn satisfies the following recursions:

$$W_{n+1} = W_n + X_{n+1}\xi_{n+1} \text{ and } T_{n+1} = T_n + mY_{n+1} + \xi_{n+1}(X_{n+1} - Y_{n+1}). \tag{12}$$

Next results will be used in the proof of Theorem 2.

Lemma 7. *Under the assumptions of Theorem 2, if $\mu_X \neq \mu_Y$, the proportion of white balls in the urn after n draws satisfies the stochastic algorithm defined by Eq. (5) where $\gamma_n = 1/T_n$,*

$$f(x) = m(\mu_Y - \mu_X)x(x-1),$$

and

$$\Delta M_{n+1} = D_{n+1} - \mathbb{E}[D_{n+1}|\mathcal{F}_n],$$

with

$$D_{n+1} = \xi_{n+1}(Z_n(Y_{n+1} - X_{n+1}) + X_{n+1}) - mZ_nY_{n+1}.$$

Proof. We check that, if $\mu_X \neq \mu_Y$, the assumptions of Definition 1 hold. Indeed,

(i) Eq. (12) shows that

$$T_n = T_0 + m\sum_{i=1}^{n} Y_i + \sum_{i=1}^{n} \xi_i(X_i - Y_i), \tag{13}$$

since the expression of T_n is similar to that in Equation (9), we have the same bound of $\gamma_n = \frac{1}{T_n}$ defined in Eq. (10).

(ii)

$$\mathbb{E}\left[(\Delta M_{n+1})^2|\mathcal{F}_n\right] \le \left(2m+m^2\right)\left(4\mu_{X^2}+\mu_{Y^2}\right)+3m^2\mu_{Y^2}+2m^2\mu_X + 2m^2\mu_X\mu_Y+4m^2(\mu_X-\mu_Y)^2 = K_\Delta.$$

(iii) $|f(Z_n)| = |m(\mu_Y-\mu_X)Z_n(Z_n-1)| \le 2m|\mu_Y-\mu_X| = K_f$,

(iv) $\mathbb{E}[\gamma_{n+1}\Delta M_{n+1}|\mathcal{F}_n] \le \frac{1}{T_n}\mathbb{E}[\Delta M_{n+1}|\mathcal{F}_n] = 0.$ □

Proposition 3. *Under the assumptions of Theorem 2, the proportion of white balls in the urn after n draws, Z_n, satisfies a.s.*

$$\lim_{n\to+\infty} Z_n = \begin{cases} 0, & \text{if } \mu_X < \mu_Y; \\ 1, & \text{if } \mu_X > \mu_Y; \\ \tilde{Z}_\infty, & \text{if } \mu_X = \mu_Y, \end{cases}$$

where $\tilde{Z}_\infty$ is a positive random variable.

Proof. Recall that, if $\mu_X \neq \mu_Y$, Z_n satisfies the stochastic algorithm of Lemma 7. As the function f is continuous, by Theorem 3 we conclude that Z_n converges $a.s.$ to the stable zero of the function h with a negative derivative, which is 1 if $\mu_X > \mu_Y$ and 0 if $\mu_X < \mu_Y$.

In the case when $\mu_X = \mu_Y$, we have $Z_{n+1} = Z_n + \frac{P_{n+1}}{T_{n+1}}$, where

$$P_{n+1} = X_{n+1}\xi_{n+1} - Z_n(mY_{n+1} + \xi_{n+1}(X_{n+1} - Y_{n+1})).$$

Since $\mathbb{E}[P_{n+1}|\mathcal{F}_n] = 0$, then Z_n is a positive martingale which converges $a.s.$ to a positive random variable $\tilde{Z}_\infty$. □

As a consequence of Proposition 3, we have

Corollary 1. *Suppose that $\mu_X \ge \mu_Y$, the total number of balls in the urn, T_n, satisfies as n tends to infinity*

$$\lim_{n\to+\infty} \frac{T_n}{n} = m\mu_X, \quad a.s.$$

Remark. The convergence in Corollary 1 holds also in $\mathbb{L}^2$.

Proof. We have

$$\frac{T_n}{n} = \frac{T_0}{n} + \frac{m}{n}\sum_{i=1}^n Y_i + \frac{1}{n}\sum_{i=1}^n \xi_i(X_i - Y_i) = \frac{T_0}{n} + \frac{m}{n}\sum_{i=1}^n Y_i + \frac{m(\mu_X-\mu_Y)}{n}\sum_{i=1}^n Z_{i-1} - \frac{G_n}{n},$$

where $G_n = \sum_{i=1}^n[\xi_i(Y_i - X_i) - \mathbb{E}(\xi_i(Y_i - X_i)|\mathcal{F}_n)]$ is the martingale difference defined in the proof of Lemma 6. Recall that G_n/n converges $a.s.$ to 0 and that Z_n converges $a.s.$ to 1 when

$\mu_X > \mu_Y$, . Then, using Cesáro lemma, we obtain the limits requested. If $\mu_X = \mu_Y$, we have $\frac{1}{n}\sum_{i=1}^{n} Y_i$ converges to μ_X. □

For the particular case when $X_n = Y_n$ for all n, we have the following results

Proposition 4 ([5]). *Let $(\Omega_l)_{l\geq 0}$ be a sequence of increasing events such that $\mathbb{P}(\cup_{l\geq 0}\Omega_l) = 1$. If there exists nonnegative Borel measurable function $\{f_l\}_{l\geq 1}$ such that for all Borel sets B*

$$\mathbb{P}\left(\Omega_l \cap W_\infty^{-1}(B)\right) = \int_B f_l(x)dx$$

then, $f = \lim_{l\to+\infty} f_l$ exists almost everywhere and f is the density of W_∞.

Lemma 8. *Define the events*

$$\Omega_l := \{W_l \geq mU \text{ and } B_l \geq mU\},$$

then, $(\Omega_l)_{l\geq 0}$ is a sequence of increasing events, moreover we have $\mathbb{P}(\cup_{l\geq 0}\Omega_l) = 1$.

Let $(p_c)_{c\in \mathrm{supp}(X)}$ the distribution of X.

Lemma 9. *For a fixed $l > 0$, there exists a positive constant κ, such that, for every $c \in supp(X)$, $n \geq l+1$, $Um \leq j \leq T_{l-1}$ and $k \leq Um(n+1)$, we have*

$$\sum_{i=0}^{m} \mathbb{P}(W_{n+1} = j+k | W_n = j+k-ci) \leq p_c\left(1 - \frac{1}{n} + \frac{\kappa}{n^2}\right). \tag{14}$$

Proof. According to Lemma 4.1 in [5], for $Um \leq j \leq T_{l-1}$, $n \geq l$ and $k \leq Um(n+1)$, the following holds:

$$\sum_{i=0}^{m}\binom{j+c(k-i)}{i}\binom{T_n - j - c(k-i)}{m-i} = \frac{T_n^m}{m!} + \frac{(1-m-2c)T_n^{m-1}}{2(m-1)!} + \cdots, \tag{15}$$

which is a polynomial in T_n of degree m with coefficients depending on W_0, B_0, m and c only.

Let $u_{n,k}(c) = \sum_{i=0}^{m} \mathbb{P}(W_{n+1} = j+k | W_n = j+k-ic)$. Applying Eq. (15) to our model we have almost surely

$$u_{n,k}(c) = p_c \sum_{i=0}^{m}\binom{j+k}{i}\binom{T_n - j - k}{m-i}\binom{T_n}{m}^{-1}$$

$$= p_c \binom{T_n}{m}^{-1}\left(\frac{T_n^m}{m!} + \frac{(1-m-2c)}{(m-1)!}T_n^{m-1} + \cdots\right) \times \left(\frac{T_n^m}{m!} + \frac{(1-m)}{2(m-1)!}T_n^{m-1} + \cdots\right)^{-1}$$

$$= p_c\left(1 - \frac{1}{n} + O\left(\frac{1}{n^2}\right)\right). \quad \square$$

4.2 Proof of Theorem 1

Recall that $(X_i)_{i\geq 1}$ (resp $(Y_i)_{i\geq 1}$) is a sequence of random variable distributed like X (resp Y). We consider the urn model evolving by the anti-diagonal matrix $Q_n = \begin{pmatrix} 0 & X_n \\ Y_n & 0 \end{pmatrix}$.

Proof of claim 1 Theorem 1**.** In order to describe the asymptotic of the urn's composition we use Lemma 6 which gives the estimate of T_n, the total number of balls in the urn after n draws. For the number of white and blue balls we have, *a.s.*

$$\frac{W_n}{n} = \frac{W_n}{T_n}\frac{T_n}{n} \text{ and } \frac{B_n}{n} = \frac{B_n}{T_n}\frac{T_n}{n},$$

using Eqs. (8), (11) and Slutsky theorem, we have almost surely, as n goes to infinity,

$$\frac{W_n}{n} \to m\sqrt{\mu_X\mu_Y}z \text{ and } \frac{B_n}{n} \to m\sqrt{\mu_X\mu_Y}(1-z).$$

These convergence hold also in $\mathbb{L}^2$.

Proof of claim 2 Theorem 1. To establish a central limit theorem, we aim to apply Theorem 3. Recall that in our model, we have $\gamma_n = 1/T_n$, then we need to find the following limits:

$$\lim_{n\to+\infty} \mathbb{E}[\left(\frac{n}{T_n}\right)^2 \Delta M_{n+1}^2 | \mathcal{F}_n] \text{ and } \lim_{n\to+\infty} \frac{n}{T_n} f'(Z_n).$$

In fact, in view of Lemma 6, we have n/T_n converges $a.s.$ to $(m\sqrt{\mu_X\mu_Y})^{-1}$ and

$$\mathbb{E}\left[(\Delta M_{n+1})^2 | \mathcal{F}_n\right] = \mathbb{E}\left[(D_{n+1})^2 | \mathcal{F}_n\right] + \mathbb{E}[D_{n+1}|\mathcal{F}_n]^2.$$

Since $\mathbb{E}[D_{n+1}|\mathcal{F}_n]^2$ converges $a.s.$ to $(f(z))^2 = 0$, we have,

$$\begin{aligned}\mathbb{E}\left((\Delta M_{n+1})^2 | \mathcal{F}_n\right] &= \mathbb{E}\left[Z_n^2(X_{n+1} - Y_{n+1})^2 - 2Z_nX_{n+1} + X_{n+1}|\mathcal{F}_n\right]\mathbb{E}\left[\xi_{n+1}^2|\mathcal{F}_n\right] + m^2\mathbb{E}\left(X^2\right) \\ &+ 2m^2\left(Z_n^2\left(\mathbb{E}(X^2\right) - \mu_X\mu_Y\right) - Z_n\mathbb{E}\left(X^2\right)\right).\end{aligned}$$

Using the fact that

$$\mathbb{E}\left[\xi_{n+1}^2|\mathcal{F}_n\right] = mZ_n(1-Z_n)\frac{T_n - m}{T_n - 1} + m^2Z_n^2$$

and that Z_n converges $a.s.$ to z, we conclude that $\mathbb{E}[D_{n+1}^2|\mathcal{F}_n]$ converges $a.s.$ to $G(z) > 0$. Applying Theorem 3, we obtain the following

$$\sqrt{n}(Z_n - z) \xrightarrow{\mathcal{L}} \mathcal{N}\left(0, \frac{G(z)}{3m^2\mu_X\mu_Y}\right).$$

Since we have

$$\frac{W_n - zT_n}{\sqrt{n}} = \sqrt{n}\left(\frac{W_n}{T_n} - z\right) \frac{T_n}{n},$$

Slutsky theorem is enough to conclude the proof.

Proof of claim 3 Theorem 1. In this particular case, the claims (1) and (2) apply and the almost sure limit of the urn's composition follows immediately as well as a central limit theorem. Furthermore, as such a case is easier, we can obtain a finer rate of convergence of the normalized number of balls in the urn. We also give another version of central limit theorem satisfied by W_n using the weak dependence between the variables $(\xi_i)_{i\geq 0}$ and the Bernstein's method.

Recall that when $Y_n = X_n$ for all $n \geq 0$, the urn is evolving according to Eq. (1) with a replacement matrix given by

$$Q_n = \begin{pmatrix} 0 & X_n \\ X_n & 0 \end{pmatrix}.$$

Theorem 1(1) applies for $z = 1/2$ and the following almost sure results follows:

$$T_n = m\mu_X n + o(n), \quad W_n = \frac{m\mu_X}{2}n + o(n) \text{ and } B_n = \frac{m\mu_X}{2}n + o(n).$$

On the other hand, the total number of balls in the urn is a sum of i.i.d. random variables $T_n = T_0 + \sum_{i=1}^n X_i$. According to the strong law of large number we get a finer rate of convergence of T_n, we have for $\delta > \frac{1}{2}$

$$T_n = m\mu_X n + o\left(\sqrt{n}\ln^\delta n\right). \tag{16}$$

Using $\frac{W_n}{n} = \frac{W_n}{T_n}\frac{T_n}{n}$ and Eq. (16), we have

$$\frac{W_n}{n} \overset{a.s}{=} (\frac{1}{2} + o(1))\left(\mu_X m + o\left(\frac{\ln^\delta n}{\sqrt{n}}\right)\right).$$

We conclude that the number of white balls in the urn after n draws, W_n, satisfies almost surely for n large enough

$$W_n = \frac{\mu_X m}{2} n + o\left(\sqrt{n}\ \ln^\delta\ n\right), \quad \delta > \frac{1}{2}.$$

Remark. In such a model, the proportion of white balls in the urn, Z_n, satisfies the stochastic approximation algorithm defined by Eq. (5) with $\gamma_n = 1/T_n$,

$$f(x) = \mu_X m(1 - 2x)$$

and

$$\Delta M_{n+1} = X_{n+1}(m - \xi_{n+1} - mZ_n) - \mu_X m(1 - 2Z_n).$$

Moreover, we propose the following result about the variance of W_n.

Proposition 5. *Under the hypothesis of Theorem 1, with $Y_n = X_n$ for all $n \geq 0$, the variance of W_n satisfies for every $\delta > \frac{1}{2}$,*

$$\mathbb{V}ar(W_n) = \frac{m\left(\sigma_X^2 + \mu_X^2\right) + m^2\sigma_X^2}{12} n + o\left(\sqrt{n}\ln^\delta n\right). \tag{17}$$

Proof. Because the number of white balls in the urn satisfies Eq. (6), we write

$$\mathbb{V}ar(W_{n+1}) = \mathbb{V}ar(W_n) + \mathbb{V}ar(X_{n+1}(m - \xi_{n+1})) + 2\,\mathbb{C}ov(W_n, X_{n+1}(m - \xi_{n+1})).$$

We have

$$\begin{aligned}
\mathbb{V}ar(X_n(m - \xi_n)) &= \mathbb{E}\left(X^2\right)\mathbb{V}ar(m - \xi_{n+1}) + \mathbb{V}ar(\mathrm{X})\mathbb{E}\left((m - \xi_{n+1})^2\right) \\
&= \left(\sigma_X^2 + \mu_X^2\right)[\mathbb{E}(\mathbb{V}ar(\xi_{n+1}|\mathcal{F}_n)) + \mathbb{V}ar(\mathbb{E}(\xi_{n+1}|\mathcal{F}_n))] + \sigma_x^2\mathbb{E}\left((m - \xi_{n+1})^2\right) \\
&= \left(\sigma_x^2 + \mu_{X^2}\right)\left(\mathbb{V}ar(mZ_n) + \mathbb{E}\left(mZ_n(1 - Z_n)\frac{T_n - m}{T_n - 1}\right)\right) + \sigma_X^2\mathbb{E}(m - \xi_n)^2.
\end{aligned} \tag{18}$$

On the other hand, since the variables $(X_i)_{i\geq 0}$ are independent then X_{n+1} and W_n are independent, thus it follows

$$\begin{aligned}
\mathbb{C}ov(W_n,\, X_{n+1}(m - \xi_{n+1})) &= \mathbb{C}ov(W_n,\, mX_{n+1}) - \mathbb{C}ov(W_n,\, X_{n+1}\xi_{n+1}) \\
&= -\mathbb{C}ov(W_n,\, X_{n+1}\xi_{n+1}) \\
&= -m\mu_X[\mathbb{E}(W_n\frac{W_n}{T_n}) + \mathbb{E}(W_n)\mathbb{E}(\frac{W_n}{T_n})] \\
&= -m\mu_X(\frac{1}{m\mu_X}(1 + o(\frac{\ln^\delta n}{\sqrt{n}}))\mathbb{V}ar(W_n))
\end{aligned} \tag{19}$$

Using Eqs. (18) and (19) and the fact that $Z_n \xrightarrow{a.s} \frac{1}{2}$ as n goes to infinity, we obtain

$$\mathbb{V}ar(W_{n+1}) = \left(1 - \frac{2}{n} + o\left(\frac{\ln^{\delta} n}{n^{\frac{3}{2}}}\right)\right)\mathbb{V}ar(W_n) + \frac{m(\sigma_X^2 + \mu_{X^2}) + m^2\sigma_X^2}{4} + o\left(\frac{\ln^{\delta} n}{\sqrt{n}}\right)$$

$$= a_n \mathbb{V}ar(W_n) + b_n,$$

where $a_n = \left(1 - \frac{2}{n} + o\left(\frac{\ln^{\delta} n}{n^{\frac{3}{2}}}\right)\right)$ and $b_n = \frac{m(\sigma_X^2 + \mu_{X^2}) + m^2\sigma_X^2}{4} + o\left(\frac{\ln^{\delta} n}{\sqrt{n}}\right)$.

Thus,

$$\mathbb{V}ar(W_n) = \left(\prod_{k=1}^{n} a_k\right)\left(\mathbb{V}ar(W_0) + \sum_{k=0}^{n-1} \frac{b_k}{\prod_{j=0}^{k} a_j}\right).$$

There exists a constant a such that $\prod_{k=1}^{n} a_k = \frac{e^a}{n^2}\left(1 + o\left(\frac{\ln^{\delta} n}{\sqrt{n}}\right)\right)$, which leads to

$$\mathbb{V}ar(W_n) = \frac{m(\sigma_X^2 + \mu_{X^2}) + m^2\sigma_X^2}{12} n + o(\sqrt{n}\ln^{\delta} n), \delta > \frac{1}{2}. \quad \square$$

In this particular case, two versions of the central limit theorem for the number of white balls are proved. The first version is deduced by Theorem 1(2) and the second one is proved using the weak dependence between the variables $(\xi_i)_{i\geq 1}$ together with Bernstein's Method.

Applying Theorem 1(2), we have $Y_n = X_n$, it follows that $\mu_Y = \mu_X$, by a simple computation for the coefficients a_i for $i \in \{0, \ldots, 4\}$ we have for $z = \frac{1}{2}$:

$$G\left(\frac{1}{2}\right) = \frac{m}{4}(\sigma_X^2 + \mu_{X^2}).$$

We conclude that, in distribution we have

$$\frac{W_n - \frac{1}{2}T_n}{\sqrt{n}} \rightarrow \mathcal{N}\left(0, \frac{m}{12}(\sigma_X^2 + \mu_{X^2})\right).$$

A second central limit theorem is satisfied by W_n. As the proof is close to that of Lemma 3 and Theorem 4 in [2], we will mention only the main steps and we refer the reader to [2] for the details. The idea of the proof is the following: Once we prove that the variables $(X_n(m - \xi_n))_{n\geq 0}$ are α-mixing variables with a strong mixing coefficient $\alpha(n) = o(\ln^{\delta} n/\sqrt{n})$, $\delta > 1/2$ (see Lemma 3 in [2] for detailed computations), Bernstein's method (see [17]) will be suitable. Consider the same notations as in Theorem 4 in [2] with

$$\tilde{\xi}_i = X_i(m - \xi_i) - \mu_X(m - \mathbb{E}(\xi_i)), \qquad S_n = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \tilde{\xi}_i$$

and N is the centered normal random variable with variance

$$\sigma^2 = \frac{m}{12}\left[(\sigma_X^2 + \mu_{X^2}) + m^2\sigma_X^2\right].$$

Actually, all that remains in this case, is to compute the variance of W_n. For that, we use Proposition 5. As a conclusion,

$$\frac{W_n - \mathbb{E}(W_n)}{\sqrt{n}} \xrightarrow{\mathcal{D}} \mathcal{N}\left(0, \frac{m}{12}(\sigma_X^2 + \mu_{X^2}) + m^2\sigma_X^2\right).$$

4.3 Proof of Theorem 2

Theorem 2 deals with unbalanced urn model with diagonal replacement matrix. We applied Proposition 1 to find the almost sure limit of the proportion of white balls in the urn. The stochastic algorithm applies only to the case when $\mu_X \neq \mu_Y$, because when $\mu_X = \mu_Y$ we fall on the case $f \equiv 0$. Furthermore, Theorem 3 does not work, in fact, by a simple computation we obtain $\sigma = 0$. Such a result is expected since that even for the case $X_n = Y_n = C$(C is constant) and $m > 1$, the fluctuations of W_n/n around its limit has not a normal distribution.

Consider the urn model defined by Eq. (1) with $Q_n = \begin{pmatrix} X_n & 0 \\ 0 & Y_n \end{pmatrix}$.

Proof of claims 1 and 2 **Theorem 2.** Corollary 1 ensures that, if $\mu_X \geq \mu_Y$ we have

$$T_n = m\mu_X n + o(n).$$

Indeed,

- If $\mu_X > \mu_Y$, we have, a.s.,

$$\lim_{n\to+\infty} \frac{W_n}{n} = \lim_{n\to+\infty} \frac{W_n}{T_n} \frac{T_n}{n} = m\mu_X.$$

Moreover, let $\tilde{G}_n = \left(\prod_{i=1}^{n-1} (1 + \frac{m\mu_Y}{T_i}) \right)^{-1} B_n$, then $(\tilde{G}_n, \mathcal{F}_n)$ is a positive martingale. There exists a positive number A such that $\prod_{i=1}^{n-1} \left(1 + \frac{m\mu_Y}{T_i}\right) \simeq An^\rho$ where $\rho = \frac{\mu_Y}{\mu_X}$. Then, as n tends to infinity we have

$$\frac{B_n}{n^\rho} \xrightarrow{a.s} B_\infty,$$

where B_∞ is a positive random variable.

- If $\mu_X = \mu_Y$, the sequences $\left(\prod_{i=1}^{n-1} (1 + \frac{m\mu_X}{T_i}) \right)^{-1} W_n$ and $\left(\prod_{i=1}^{n-1} (1 + \frac{m\mu_Y}{T_i}) \right)^{-1} B_n$ are $\mathcal{F}_n$ -martingales such that $\left(\prod_{i=1}^{n-1} (1 + \frac{m\mu_X}{T_i}) \right)^{-1} \simeq Bn$, where $B > 0$, then, as n tends to infinity, we have

$$\frac{W_n}{n} \to W_\infty \text{ and } \frac{B_n}{n} \to \tilde{B}_\infty,\ a.s.,$$

where W_∞ and $\tilde{B}_\infty$ are positive random variables satisfying $\tilde{B}_\infty = m\mu_X - W_\infty$.

Proof of claim 3 **Theorem 2.** We consider the case when $Y_n = X_n$ for all $n \geq 0$, The urn model is then evolving according to the recursive Eq. (1) with the replacement matrix

$$Q_n = \begin{pmatrix} X_n & 0 \\ 0 & X_n \end{pmatrix}.$$

Since Theorem 2(2) applies to that case, we obtain the following strong law of large number

$$\frac{W_n}{n} \xrightarrow{a.s} W_\infty \text{ and } \frac{B_n}{n} \xrightarrow{a.s} (\mu_X m - W_\infty),$$

where W_∞ is a positive random variable. Furthermore, as T_n is a sum of i.i.d. random variables then T_n satisfies for every $\delta > \frac{1}{2}$

$$T_n \overset{a.s}{=} \frac{\mu_X m}{2} n + o\left(\sqrt{n} \ln^\delta n\right), \quad a.s. \tag{20}$$

To prove that W_∞ is absolutely continuous, we follow the proof of Theorem 4.2 in [5] and we give the main steps. The idea is the following: given the sequence of increasing event Ω_l

defined in Lemma 8, if we show that the restriction of W_∞ on every $\Omega_{l,j} = \{\omega;\ W_l(\omega) = j\}$ has a density for each j, with $Um \leq j \leq T_{l-1}$, then Proposition 4 ensures the existence of the density of W_∞ almost every where. In fact, for a fixed l and $n \geq l+1$, we denote by $v_{n,j} = \max_{0\leq k\leq Umn} \mathbb{P}(W_{l+n} = j+k | W_l = j)$. We have the following inequality:

$$v_{n+1,j} \leq \max_{0\leq k\leq Um(n+1)} \left\{ \sum_{i=0}^{m} \sum_{c\in \mathrm{supp}(X)} \mathbb{P}(W_{l+n+1} = j+k | W_{l+n} = j+k-ci) \right\}$$

$$\leq \max_{0\leq k\leq Um(n+1)} \left\{ \sum_{i=0}^{m} \sum_{c\in \mathrm{supp}(X)} \mathbb{P}(W_{l+n+1} = j+k | W_{l+n} = j+k-ci) \times \mathbb{P}(W_{l+n} = j+k-ci | W_l = j) \right\}$$

$$\leq \max_{0\leq k\leq Um(n+1)} \sum_{i=0}^{m} \sum_{c\in \mathrm{supp}(X)} \mathbb{P}(W_{l+n+1} = j+k | W_{l+n} = j+k-ci)$$

$$\times \max_{0\leq \tilde{k}\leq Umn} \mathbb{P}(W_{l+n} = j+\tilde{k} | W_l = j) \leq \sum_{c\in \mathrm{supp}(X)} p_c \left(1 - \frac{1}{n+l} + \frac{\kappa}{(n+l)^2}\right) v_{n,j}$$

$$= \left(1 - \frac{1}{n+l} + \frac{\kappa}{(n+l)^2}\right) v_{n,j}.$$

This implies that there exists some positive constant $C(l)$, depending on l only, such that, for a fixed l and for all $n \geq l+1$, we get

$$\max_{0\leq k\leq m(n-l)} \mathbb{P}(W_n = j+k | W_l = j) \leq \prod_{i=l}^{n} \left(1 - \frac{1}{i} + \frac{\kappa}{i^2}\right) \leq \frac{C(l)}{n}. \tag{21}$$

Let $\varepsilon > 0$ and $\delta = \frac{\varepsilon}{C(l)}$, and setting $x_1 < x_1' \leq x_2 < x_2' \leq \ldots \leq x_r < x_r'$ such that $\sum_{i=1}^{r} |x_i' - x_i| \leq \delta$. By Fatou's lemma we have

$$\sum_{i=1}^{r} \mathbb{P}(\{x_i \leq W_\infty \leq x'\} \cap \Omega_{l,\ j}) \leq \sum_{i=1}^{r} \liminf \mathbb{P}\left(x_i \leq \frac{W_n}{n} \leq x_i' | W_l = j\right) \mathbb{P}(\Omega_{l,\ j})$$

$$\leq \sum_{i=1}^{r} \liminf \left(((x_i' - x_i)n + 1) \frac{C(l)}{n} \right)$$

$$\leq \sum_{i=1}^{r} (x_i' - x_i) C(l) = \varepsilon.$$

Then the proof follows.

Outlook: *We suggest that if we replace the boundedness hypothesis of the variablesX andY by the assumption that X and Y have finite moments of order 2, our results remain true.*

References

[1] R. Aguech, Limit theorems for random triangular urns schemes, J. Appl. Probab. 46 (3) (2009) 827–843.

[2] R. Aguech, et al., A generalized urn model with multiple drawing and random addition, Ann. Inst. Statist. Math. (2018) 1–20.

[3] K.B. Athrea, S. Karlin, Embedding of urns schemes into continuous time Markov branching processes and related limit theorems, Ann. Math. Stat. 39 (1968) 1801–1817.

[4] M.R. Chen, M. Kuba, On generalized polya urn models, J. Appl. Probab. 50 (4) (2013) 909–1216.

[5] M.R. Chen, C.Z. Wei, A new urn model, J. Appl. Probab. 42 (4) (2005) 964–976.

[6] M. Duflo, Random Iterative Models, Springer-Verlag, Berlin, 1997.

[7] F. Eggenberger, G. Pólya, Über die statistik verkeletter vorgange, Z. Angew. Math. Mech. 3 (1923) 279–289.

[8] R.N. Goldman, Polya's urn model and computer aided geometric design, SIAM J. Algebr. Discrete Methods 6 (1) (1985) 1–28.

[9] S. Janson, Functionnal limt theorems for multitype branching processes and generalized pólya urns, Stochastic Process. Appl. 110 (2) (2004) 177–245.

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

[11] N.L. Johnson, S. Kotz, Urn Models and Their Application, John Wiley & Sons, 1977.

[12] M. Kuba, H. Mahmoud, Two-colour balanced affine urn models with multiple drawings II: large-index and triangular urns, 2016, arXiv:150909053.

[13] M. Kuba, H. Mahmoud, Two-colour balanced affine urn models with multiple drawings I: Central limit theorem, Adv. Appl. Probab. 90 (2017) 1–26.

[14] M. Kuba, H. Sulzbach, On martingale tail sums in affine two-color urn models with multiple drawings, J. Appl. Probab. 54 (2017) 1–21.

[15] M. Kuba, et al., Analysis of a generalized Friedman's urn with multiple drawings, Discrete Appl. Math. 161 (18) (2013) 2968–2984.

[16] N. Lasmar, et al., Multiple drawing multi-colour urns by stochastic approximation, J. Appl. Probab. 55 (1) (2018) 254–281.

[17] Z. Lin, C. Lu, Limit Theory for Mixing Dependent Random Variables, Kluwer Academic Publishers, Boston, 1996.

[18] H. Mahmoud, Random spouts as internet model and pólya processes, Acta inform. 41 (1) (2004) 1–18.

[19] G. Pagés, S. Laruelle, Randomized urns models revisited using stochastic approximation, Ann. Appl. Probab. 23 (4) (2013) 1409–1436.

[20] H. Renlund, Generalized polya urns via stochastic approximation, 2010, arxiv:10023716v1.

[21] H. Renlund, Limit theorem for Stochastic approximation algorithm, 2011, arxiv:11024741v1.

[22] L.J. Wei, An application of an urn model to the design of sequential controlled clinical trials, J. Am. Stat. Assoc. 73 (1978) 559–563.

Corresponding author
Aguech Rafik can be contacted at: raguech@ksu.edu.sa