Unbalanced multi-drawing urn with random addition matrix

Rafik Aguech*[1] and Olfa Selmi[2]

[1]Departement of Statistics and operation research, King Saoudian University Erryadh K.S.A.
[1] Departement of mathematics Faculté des sciences de Monastir Monatir, Tunisia.
[2]Departement of Mathematics Faculté des Sciences de Monastir Monastir, Tunisia. †

Abstract

In this paper, we consider a two color multi-drawing urn model. At each discrete time step, we draw a sample of m balls ($m \geq 2$), which will be returned to the urn together with a random number of balls. The replacement rule is a 2×2 matrix depending on X and Y, two discrete positive random variables with finite means and variances. Using a stochastic approximation algorithm, we study the asymptotic behavior of the urn.
Keywords: central limit theorem, unbalanced urn, martingale, stochastic algorithm.

1 Introduction

The classical Pólya urn was introduced by Pólya and Eggenberger [2] 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.

Afterward, there were many generalisations and urn model become a simple tool to describe several models such finance, clinical trials (see [3], [8]), biology (see [16]), computer sciences, internet (see [6],[19]), etc.
Recently, H. Mahmoud, M.R. Chen, C.Z Wei, M. kuba and H. Sulzbach [10, 11, 12, 13, 14, 15], have focused on the multidrawing urn. Instead of picking a ball, one picks a sample of m balls ($m \geq 1$), say l white and $(m - l)$ blue balls. The pick is returned back to the urn together with a_{m-l} white and b_l blue balls, where a_l and $b_l, 0 \leq l \leq m$ are integers. At first, they treated two particular cases when $\{a_{m-l} = c \times l \quad \text{and} \quad b_{m-l} = c \times (m - l)\}$ and when $\{a_{m-l} = c \times (m - l) \quad \text{and} \quad b_{m-l} = c \times l\}$, where c is a positive constant. By different methods as martingales and moment methods, the 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 C. Mailler, N. Lasmer and S. Olfa [1], they generalized this model and looked at the case of more than two colors.

In the literature, there are few works about unbalanced urns, mainly the papers of R. Aguech [17] and S. Janson [20] who dealt with urns with a simple pick. Whereas H. Renlund [4, 5] considered a model with two draws at each epoch. The authors used different models such as martingale, stochastic algorithm and embedding in continuous time. They gave limit theorems of the asymptotic behavior of the urn.
In this paper, we deal with a two color unbalanced urn class with multiple drawing and random addition. Let denote by W_n (resp B_n) the number of white balls (resp blue balls) and by T_n the total number of balls in the urn at time n. Let m a non null integer, the model we study is defined as follows: At each

*Corresponding author: rafik.aguech@ipeit.rnu.tn
†selmiolfa3@yahoo.fr

discrete step we draw a sample of m balls (we assume that the initial composition of the urn is more than m to ensure that the first draw is possible). Let ξ_n be the number of white balls among the n^{th} sample, then 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, where $(X_n)_{n\geq 1}$ (resp $(Y_n)_{n\geq 1}$) are independent copies of X (resp Y), two positive independent random variables with finite means and variances.
We summarize the evolution of the urn by the following recurrence

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

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

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

We aim to extend the results of recent works [10], [11] and [18] where the authors characterized the urn models defined by Equation 1 for the following cases $Q_n = \begin{pmatrix} a & 0 \\ 0 & a \end{pmatrix}$, $Q_n = \begin{pmatrix} 0 & a \\ a & 0 \end{pmatrix}$, $Q_n = \begin{pmatrix} a & 0 \\ 0 & b \end{pmatrix}$ and $Q_n = \begin{pmatrix} 0 & a \\ b & 0 \end{pmatrix}$, where a, b are strictly positive integers. Before each draw n we replace a and b respectively by X_n and Y_n. 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. In the subsection3.1 we reformulate Theorem 1 in [4] to make it suitable to our model. The proofs of the main results are detailed in subsection 3.2.

Notation: For a random variable R, we denote by $\mu_R = \mathbb{E}(R)$ and $\sigma_R^2 = \mathbb{V}ar(R)$. Note that μ_X, μ_Y, σ_X^2 and σ_Y^2 are finite.
For x_n and y_n two sequences of real numbers such that $y_n \neq 0$ for all n, we say that x_n is a little-o of y_n if and only if $\lim\limits_{n\to+\infty} \frac{x_n}{y_n} = 0$. We then denote $x_n = o(y_n)$.

2 Main Results

In the present section we give the main results of this worksheet. As mentioned in the introduction, we study urn models evolving according to the recursion of Equation (1). Recall that in the following we consider two sequences of random random variables $(X_n)_{n\geq 1}$ (resp $(Y_n)_{n\geq 1}$) iid copies of X (resp Y), where X and Y are strictly positive random variables with finite means and variances. We show in the following that the use of the stochastic algorithms was fruitful and prove that the normalized number of balls in the urn converges almost surely to a finite limits, furthermore, whenever the matrix Q_n is anti-diagonal, we show that the fluctuation of W_n around its limit is normal. In the case when Q_n is diagonal we are not able yet to characterize these fluctuations.

Theorem 1. *Consider the urn model evolving by the matrix* $Q_n = \begin{pmatrix} 0 & X_n \\ X_n & 0 \end{pmatrix}$. *Let* $\delta > \frac{1}{2}$, *we have the following results*

1. *The total number of balls in the urn after n draws satisfies for all* $\delta > \frac{1}{2}$

$$T_n \stackrel{a.s}{=} m\mu_X n + o(\sqrt{n}\ \ln(n)^\delta), \tag{2}$$

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

$$W_n \stackrel{a.s}{=} \frac{m\mu_X}{2} n + o(\sqrt{n}\ \ln(n)^\delta), \tag{3}$$

and

$$B_n \stackrel{a.s}{=} \frac{m\mu_X}{2} n + o(\sqrt{n}\ \ln(n)^\delta). \tag{4}$$

The following results deal with two versions of central limit theorem of W_n. *In distribution we have*

2. *Let* $\Sigma = \frac{m(\sigma_X^2+\mu_X^2)}{12}$, *then,*
$$\lim_{n\to+\infty} \frac{W_n - \frac{1}{2}T_n}{\Sigma\sqrt{n}} = \mathcal{N}\Big(0,1\Big). \tag{5}$$

3. *Let* $\Sigma_1 = \frac{m(\sigma_X^2+\mu_X^2)+m^2\sigma_X^2}{12}$, *then,*
$$\lim_{n\to+\infty} \frac{W_n - \mathbb{E}(W_n)}{\Sigma_1\sqrt{n}} = \mathcal{N}\Big(0,1\Big). \tag{6}$$

In the following we show that, unlike the previous theorem, the stochastic algorithm's theorem does not work since we fall in the case when the function $f \equiv 0$. We are able only to give the almost sure limit of the urn's composition. These results are not surprising considering that even in the deterministic case $X_n = c$ (c is constant), then fluctuation of W_n around its limit still unknown.

Theorem 2. *Consider the urn model evolving by the matrix* $Q_n = \begin{pmatrix} X_n & 0 \\ 0 & X_n \end{pmatrix}$. *The total number of balls in the urn after* n *draws satisfies for all* $\delta > \frac{1}{2}$,
$$T_n \overset{a.s}{=} m\mu_X n + o(\sqrt{n}\ \ln(n)^\delta). \tag{7}$$
Furthermore, there exists a positive random variable $\tilde{W}_\infty$, *such that the composition of the urn satisfies*
$$W_n \overset{a.s}{=} \tilde{W}_\infty n + o(n), \tag{8}$$
and
$$B_n \overset{a.s}{=} (m\mu_X - \tilde{W}_\infty)n + o(n). \tag{9}$$

Remark: The random variable $\tilde{W}_\infty$ is absolutely continuous whenever X is bounded.
In the next theorem, the stochastic approximation's theorem was successful to determine the almost sure limit of each color and to prove a central limit theorem satisfied by W_n.

Theorem 3. *Consider the urn model evolving by the matrix* $Q_n = \begin{pmatrix} 0 & X_n \\ Y_n & 0 \end{pmatrix}$. *Let* $z := \frac{\sqrt{\mu_X}}{\sqrt{\mu_X}+\sqrt{\mu_Y}}$, *we have the following results:*

1. *The total number of balls in the urn after* n *draws satisfies*
$$T_n \overset{a.s}{=} m\sqrt{\mu_X}\sqrt{\mu_Y}\ n + o(n), \tag{10}$$
and the number of white and blue balls in the urn after n *draws satisfy almost surely*
$$W_n \overset{a.s}{=} m\sqrt{\mu_X}\sqrt{\mu_Y}\ z\ n + o(n), \tag{11}$$
and
$$B_n \overset{a.s}{=} m\sqrt{\mu_X}\sqrt{\mu_Y}(1-z)\ n + o(n). \tag{12}$$

2. *Furthermore, the normalized number of white balls in the urn satisfy the following central limit theorem*
$$\frac{W_n - zT_n}{\sqrt{n}} \xrightarrow{\mathcal{L}} \mathcal{N}\Big(0, \frac{G(z)}{3}\Big), \tag{13}$$
where,
$$G(x) = \sum_{i=0}^{4} a_i x^i,$$
with
$a_0 = m^2(\sigma_X^2 + \mu_X^2)$,
$a_1 = m(1-2m)(\sigma_X^2 + \mu_X^2)$,
$a_2 = 3m(m-1)(\sigma_X^2 + \mu_X^2) - 2m(m-1)\mu_X\mu_Y$,
$a_3 = m\mathbb{E}(X-Y)^2 - 2(m^2-m)\big(\sigma_X^2 + \mu_X^2 - \mu_X\mu_Y\big)$,
and $a_4 = m(m-1)\mathbb{E}(X-Y)^2$.

Theorem 4. *Consider the urn evolving by the matrix* $Q_n = \begin{pmatrix} X_n & 0 \\ 0 & Y_n \end{pmatrix}$. *We have the following results:*

1. *If* $\mu_X > \mu_Y$, *let* $\rho = \frac{\mu_Y}{\mu_X}$,
   The total number of balls in the urn after n *draws satisfies almost surely*
$$T_n = m\mu_X n + o(n), \tag{14}$$
   Furthermore,
$$W_n \overset{a.s}{=} m\mu_X n + o(n) \tag{15}$$
   and
$$B_n \overset{a.s}{=} B_\infty n^\rho + o(n^\rho), \tag{16}$$
   where B_∞ *is a positive random variable.*

2. *If* $\mu_X = \mu_Y$, *The composition of the urn after* n *draws satisfies almost surely*
$$T_n = m\mu_X n + o(n) \tag{17}$$
   ,
$$W_n = W_\infty n + o(n) \tag{18}$$
   and
$$B_n = (\mu_X m - W_\infty)\ n + o(n), \tag{19}$$
   where W_∞ *is a positive random variable.*

Remark: The case when $\mu_X < \mu_Y$ is obtained by interchanging the colors.
Example: Let $m = 1$, this particular case was studied by R. Aguech [17]. Using martingales and branching processes, R. Aguech proved the following results:
If $\mu_X > \mu_Y$,
$$W_n = \mu_X n + o(n), \quad B_n = Dn^\rho \quad \text{and} \quad T_n = \mu_X n + o(n),$$
where $\rho = \frac{\mu_Y}{\mu_X}$ and D is a positive random variable.
If $\mu_X = \mu_Y$,
$$W_n = \mu_X \frac{W}{W+B} n + o(n) \quad \text{and} \quad B_n = \mu_X \frac{B}{W+B} n + o(n),$$
where W and B are positive random variables obtained by embedding some martingales in continuous time.

3 Proofs

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 [9] for example), we adapt the version of H. Renlund in [4, 5].

3.1 A basic tool: Stochastic approximation

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*
$$U_{n+1} - U_n = \gamma_{n+1}\big(f(U_n) + \Delta M_{n+1}\big), \tag{20}$$
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* $[0, 1]$ *onto* $\mathbb{R}$ *and the following conditions hold almost surely.*

(i) $\frac{c_l}{n} \leq \gamma_n \leq \frac{c_u}{n}$,

(ii) $\mathbf{E}(\Delta M_{n+1}^2 | \mathcal{F}_n) \leq K_u$,

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

(iv) $\mathbb{E}[\gamma_{n+1}\Delta M_{n+1}|\mathcal{F}_n] \leq K_e\gamma_n^2$,

where the constants c_l, c_u, K_u, K_f, and K_e are positive real numbers.

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

Theorem 5. *Let U_n be a stochastic algorithm defined by Equation (20). If f is continuous, then $\lim\limits_{n\to+\infty} U_n$ exists almost surely and is in Q_f. Furthermore, if p is a stable zero, then $\mathbb{P}\Big(U_n \longrightarrow p\Big) > 0$.*

Remark: Theorem 5 is close to Theorem 1 in [4]. Trying to apply Renlund's theorem we check that assumption $(\tilde{ii})$ namely $|\Delta M_n| \leq K_u$ cannot be satisfied in our case only if the variables X and Y are bounded which is not always true. Thus we think to replace that assumption by another satisfied by such a model while keeping the same conclusion (i.e the almost sure convergence to a stable zero of the function f).

Proof of Theorem 5. The proof is close to Theorem 1 in [4], for the convenience of the reader, we resume the proof and we mention the main steps. In fact, the following lemmas are useful.

Lemma 1. *Let $V_n = \sum_{i=1}^n \gamma_i\Delta M_i$. Then, V_n converges almost surely.*

Proof. Set $A_i = \gamma_i\Delta M_i$ and $\tilde{A}_i = \mathbb{E}[A_i|\mathcal{F}_{i-1}]$. Define the martingale $C_n = \sum_{i=1}^n (A_i - \tilde{A}_i)$, then

$$\begin{aligned}\mathbb{E}(C_n^2) &\leq \sum_{i=1}^n \mathbb{E}(A_i^2) = \sum_{i=1}^n \mathbb{E}(\gamma_i^2\Delta M_i^2)\\ &\leq \sum_{i=1}^n \frac{c_u^2}{i^2}\mathbb{E}(\Delta M_i^2),\end{aligned}$$

if there exists some positive constant K_u such that $\mathbb{E}[\Delta M_{n+1}^2|\mathcal{F}_n] \leq K_u$, we conclude that C_n is an L^2- martingale and thus converges almost surely.
Next, since

$$\begin{aligned}\sum_{i\geq 1}|\tilde{A}_i| &\leq \sum_{i\geq 1}\mathbb{E}(\gamma_i\Delta M_i|\mathcal{F}_{i-1})\\ &\leq \sum_{i\geq 1} K_e\gamma_{i-1}^2\\ &\leq K_e c_u^2\sum_{i\geq 1}\frac{1}{(i-1)^2} < \infty,\end{aligned}$$

the series $\sum_{i\geq 1} A_i$ must also converges almost surely. ☐

Lemma 2. *Let $U_\infty = \bigcap_{n\geq 1}\overline{\{U_n, U_{n+1}, ...\}}$ be the set of accumulation point of $\{U_n\}$ and $Q_f = \{x; f(x) = 0\}$ be the zeros of f. Suppose f is continuous. Then,*

$$\mathbb{P}\Big(U_\infty \subseteq Q_f\Big) = 1.$$

Proof. See [4] ☐

Next, we prove the main result of Theorem 5 . If $\lim\limits_{n\to+\infty} U_n$ does not exist, we can find two rational numbers in the open interval $]\liminf\limits_{n\to+\infty} U_n, \limsup\limits_{n\to+\infty} U_n[$.
Let $p < q$ 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 theorem 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 Q_f.

Case 1: p **and** q **are not in the same connected component of** Q_f**.**
See the proof in [4].

Case 2: p **and** q **are in the same connected component of** Q_f**.**
Let p and q be two arbitrary rational numbers such that p and q are in the same connected component of Q_f. Assume that $\liminf_{n\to+\infty} U_n \leq p$ and fix an arbitrary ε such a way that $0 \leq \varepsilon \leq q - p$.
We aim to show that $\limsup_{n\to+\infty} U_n \leq q$ i.e, it is sufficient to show that $\limsup_{n\to+\infty} U_n \leq p + \varepsilon$.
In view of Lemma 1, we have $V_n = \sum_{i=1}^n \gamma_i \Delta M_i$ converges a.s, then, there exists a stochastic $N_1 > 0$ such that for $n, m > N_1$ we have $|V_n - V_m| < \frac{\varepsilon}{4}$ and $\gamma_n \Delta M_n \leq \frac{\varepsilon}{4}$.
Let $N = max(\frac{4K_f}{\varepsilon}, N_1)$. By assumption, there is some stochastic $n > N$ such that $U_n - p < \frac{\varepsilon}{2}$.
Let

$$\tau_1 = \inf\{k \geq n; U_k \geq p\} \quad \text{and} \quad \sigma_1 = \inf\{k > \tau_1; U_k < p\},$$

and define, for $n \geq 1$,

$$\tau_{n+1} = \inf\{k > \sigma_n; U_k \geq p\} \quad \sigma_{n+1} = \inf\{k > \tau_n; U_k < p\}.$$

For all k we have

$$U_{\tau_k} = U_{\tau_k - 1} + \gamma_{\tau_k}(f(U_{\tau_k - 1}) + \Delta M_{\tau_k}).$$

Recall that $\gamma_{\tau_k} f(U_{\tau_k - 1}) \leq \frac{K_f}{\tau_k - 1} \leq \frac{K_f}{n}$ and for $n \geq N \geq \frac{4K_f}{\varepsilon}$ we have $\gamma_{\tau_k} f(X_{\tau_k - 1}) < \frac{\varepsilon}{4}$. It follows,

$$\gamma_{\tau_k}(f(U_{\tau_k - 1}) + \Delta M_{\tau_k}) \leq \frac{K_f}{n} + \frac{\varepsilon}{4} \leq \frac{\varepsilon}{4} + \frac{\varepsilon}{4} = \frac{\varepsilon}{2}.$$

Note that $f(x) = 0$ when $x \in [p, q]$ (p and q are in Q_f). For j such that $\tau_k + j - 1$ is a time before the exit time of the interval $[p, q]$ (i.e $U_{\tau_k + j - 1} \in [p, q]$), we have

$$U_{\tau_k + j} = U_{\tau_k} + V_{\tau_k + j} - V_{\tau_k}.$$

As $|V_{\tau_k + j} - V_{\tau_k}| < \frac{\varepsilon}{4}$, we have $U_{\tau_k + j} \leq p + \frac{\varepsilon}{2} + \frac{\varepsilon}{4} \leq p + \varepsilon$, the process will never exceed $p + \varepsilon$ before the time σ_{k+1}. We conclude that $\sup_{k \geq n} U_k \leq p + \varepsilon$.
To establish that the almost sure limit of U_n is among the stable point set, we refer the reader to [4] to see a detailed proof.

$\square$

Theorem 6 (see [5]). *Let* $(U_n)_{n\geq 0}$ *satisfying Equation* (20) *such that* $\lim_{n\to+\infty} U_n = U^\star$*. Let* $\hat{\gamma}_n := n\gamma_n \hat{f}(U_{n-1})$ *where* $\hat{f}(x) = \frac{-f(x)}{x - U^\star}$*. Assume that* $\hat{\gamma}_n$ *converges almost surely to some limit* $\hat{\gamma}$*. Then, if* $\hat{\gamma} > \frac{1}{2}$ *and if* $\mathbb{E}[(n\gamma_n \Delta M_n)^2 | \mathcal{F}_{n-1}] \to \sigma^2 > 0$*, we have as n tends to infinity*

$$\sqrt{n}(U_n - U^\star) \xrightarrow{\mathcal{D}} \mathcal{N}\Big(0, \frac{\sigma^2}{2\hat{\gamma} - 1}\Big).$$

3.2 Proof of the main results

Proof of Theorem 1. Consider the urn model defined by Equation (1) with $Q_n = \begin{pmatrix} 0 & X_n \\ X_n & 0 \end{pmatrix}$. We have the following recursions:

$$W_{n+1} = W_n + X_{n+1}(m - \xi_{n+1}) \quad \text{and} \quad T_{n+1} = T_n + mX_{n+1}. \tag{21}$$

Proof of claim 1

Lemma 3. *Let $Z_n = \frac{W_n}{T_n}$ be the proportion of white balls in the urn after n draws. Then, Z_n satisfies the stochastic approximation algorithm defined by Equation(20) with $\gamma_n = \frac{1}{T_n}$, $f(x) = \mu_X m(1-2x)$ and $\Delta M_{n+1} = X_{n+1}(m - \xi_{n+1} - mZ_n) - \mu_X m(1-2Z_n)$.*

Proof. We need to check the conditions of definition 1.

(i) Recall that $T_n = T_0 + m\sum_{i=1}^n X_i$, with $(X_i)_{i\geq 1}$ are iid random variables. It follows, by Rajechman strong law of large numbers, that

$$T_n \overset{a.s}{=} \mu_X mn + o(\sqrt{n}\ \ln(n)^\delta), \quad \delta > \frac{1}{2}, \tag{22}$$

and we conclude that $\frac{1}{(m\mu_X+1)n} \leq \frac{1}{T_n} \leq \frac{2}{m\mu_X n}$, let $c_l = \frac{1}{m\mu_X+1}$ and $c_u = \frac{2}{m\mu_X}$,

(ii) $\mathbb{E}[\Delta M_{n+1}^2|\mathcal{F}_n] \leq (6m^2+m)\mathbb{E}(X^2) + 9m^2\mu_X^2 = K_u$,

(iii) $|f(Z_n)| = m\mu_X|1-2Z_n| \leq 3m\mu_X = K_f$,

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

□

Proposition 1. *The proportion of white balls in the urn after n draws, Z_n, converges almost surely to $\frac{1}{2}$.*

Proof of Proposition 1. Since the process Z_n satisfies the stochastic approximation algorithm defined by Equation (20), we apply Theorem 5. As the function f is continuous we conclude that Z_n converges almost surely to $\frac{1}{2}$: the unique stable zero of the function f. □

We apply the previous results to the urn composition. As we can write $\frac{W_n}{n} = \frac{W_n}{T_n}\frac{T_n}{n}$, we deduce from Proposition 1 and Equation (22) that $\frac{W_n}{n} \overset{a.s}{=} \left(\frac{1}{2} + o(1)\right)\left(\mu_X m + o\left(\frac{\ln(n)^\delta}{\sqrt{n}}\right)\right)$, then this corollary follows:

Corollary 1. *The number of white balls in the urn after n draws, W_n, satisfies for n large enough*

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

Proof of claim 2 We aim to apply Theorem 6. For this reason, we need to find these limits:

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

We have

$$\begin{aligned} \mathbb{E}[\Delta M_{n+1}^2|\mathcal{F}_n] &= \mathbb{E}(X_{n+1}^2)\mathbb{E}[(m-\xi_{n+1}-mZ_n)^2|\mathcal{F}_n]) + \mu_X^2\mathbb{E}[(m-2mZ_n)^2|\mathcal{F}_n] \\ &\quad -2\mu_X^2\mathbb{E}[(m-\xi_{n+1}-mZ_n)(m-2mZ_n)|\mathcal{F}_n] \\ &= (\sigma_X^2+\mu_X^2)\Big[m^2 - 4m^2Z_n + 4m^2Z_n^2 + mZ_n(1-Z_n)\frac{T_n-m}{T_n-1}\Big] - \mu_X^2[m^2+4m^2Z_n^2-4m^2Z_n]. \end{aligned}$$

As n tends to infinity, we have $Z_n \xrightarrow{a.s} \frac{1}{2}$ and $\frac{T_n-m}{T_n-1} \xrightarrow{a.s} 1$. Then,

$$\lim_{n\to\infty} \mathbb{E}[\Delta M_{n+1}^2|\mathcal{F}_n] \overset{a.s}{=} (\sigma_X^2+\mu_X^2)\frac{m}{4} \quad \text{and} \quad \lim_{n\to\infty} -\frac{n}{T_n}f'(Z_n) \overset{a.s}{=} 2.$$

According to Theorem 6, $\sqrt{n}(Z_n - \frac{1}{2})$ converges in distribution to $\mathcal{N}(0, \frac{\sigma_X^2+\mu_X^2}{12\mu_X^2 m})$. Finally, by writing $\left(\frac{W_n - \frac{1}{2}T_n}{\sqrt{n}}\right) = \sqrt{n}(Z_n - \frac{1}{2})\frac{T_n}{n}$, we conclude using Slutsky theorem.

Proof of claim 3 We are proving a second version of central limit theorem satisfied by W_n. as the proof is close to that of Lemma 3 and Theorem 4 in [18], we will mention only the main steps and we refer the reader to [18] 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 $\alpha(n) = o\Big(\frac{\ln(n)^\delta}{\sqrt{n}}\Big)$, $\delta > \frac{1}{2}$ (see Lemma 3 in [18] for detailed computations), the Bernstein method (see [21]) will be perfectly applied. Consider the same notations as in Theorem 4 in [18] with $\tilde{\xi_i} = X_i(m-\xi_i) - \mu_X(m - \mathbb{E}(\xi_i))$, $S_n = \frac{1}{\sqrt{n}}\sum_{i=1}^n \tilde{\xi_i}$ and N is the centered normal random variable with variance $\sigma^2 = \frac{m(\sigma_X^2+\mu_X^2)+m^2\sigma_X^2}{12}$. All that we need more in that case is to compute the variance of W_n. In fact we have the following,

Proposition 2. *The variance of W_n satisfies*

$$\mathbb{V}ar(W_n) = \frac{m(\sigma_X^2+\mu_X^2)+m^2\sigma_X^2}{12}\, n + o(\sqrt{n}\ \ln(n)^\delta), \quad \delta > \frac{1}{2}. \tag{23}$$

Proof of Proposition 2. Recall that the number of white balls in the urn satisfies Equation (21), then

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

We have $\mathbb{V}ar(X_n(m-\xi_n)) = (\sigma_X^2+\mu_X^2)\Big(\mathbb{V}ar(mZ_{n-1}) + \mathbb{E}\Big(mZ_{n-1}(1-Z_{n-1})\frac{T_{n-1}-m}{T_{n-1}-1}\Big)\Big) + \sigma_X^2\mathbb{E}(m-\xi_n)^2$.

Using Equation (22) and the fact that $Z_n \overset{a.s}{\to} \frac{1}{2}$, we obtain

$$\begin{aligned}\mathbb{V}ar(W_{n+1}) &= \Big(1 - \frac{2}{n} + o\Big(\frac{\ln(n)^\delta}{n^{\frac{3}{2}}}\Big)\Big)\mathbb{V}ar(W_n) + \frac{m(\sigma_X^2+\mu_X^2)+m^2\sigma_X^2}{4} + o\Big(\frac{\ln(n)^\delta}{\sqrt{n}}\Big)\\ &= a_n\mathbb{V}ar(W_n) + b_n,\end{aligned}$$

where $a_n = \Big(1-\frac{2}{n} + o\Big(\frac{\ln(n)^\delta}{n^{\frac{3}{2}}}\Big)\Big)$ and $b_n = \frac{m(\sigma_X^2+\mu_X^2)+m^2\sigma_X^2}{4} + o\Big(\frac{\ln(n)^\delta}{\sqrt{n}}\Big)$.
Thus,

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

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

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

□

As a conclusion, $\frac{W_n - \mathbb{E}(W_n)}{\sqrt{n}}$ converges in distribution to the variable N. □

Proof of Theorem 2. Consider the urn model defined by Equation (1) with $Q_n = \begin{pmatrix} X_n & 0 \\ 0 & X_n\end{pmatrix}$. The following recurrences hold:

$$W_{n+1} = W_n + X_{n+1}\xi_{n+1} \quad \text{and} \quad T_{n+1} = T_n + mX_{n+1}. \tag{24}$$

As T_n is a sum of iid random variables then T_n satisfies the following

$$T_n \overset{a.s}{=} \frac{\mu_X m}{2}n + o(\sqrt{n}\ln(n)^\delta); \quad \delta > \frac{1}{2}. \tag{25}$$

The processes $\tilde{M}_n = \prod_{k=1}^{n-1}\Big(\frac{T_k}{T_k+m\mu_X}\Big)W_n$ and $\tilde{N}_n = \prod_{k=1}^{n-1}\Big(\frac{T_k}{T_k+m\mu_X}\Big)B_n$ are two $\mathcal{F}_n$ positive martingales. In view of Equation (25), there exists a positive constant γ; $\prod_{k=1}^{n-1}\Big(\frac{T_k}{T_k+m\mu_X}\Big) \overset{a.s}{=} \dfrac{e^\gamma}{n}\Big(1 + o\Big(\dfrac{\ln(n)^\delta}{\sqrt{n}}\Big)\Big)$ for $\delta > \frac{1}{2}$. Thus, there exist nonnegative random variables $\tilde{W}_\infty$ and $\tilde{B}_\infty$ such that $\tilde{W}_\infty + \tilde{B}_\infty \overset{a.s}{=} m\mu_X$ and

$$\frac{W_n}{n} \xrightarrow{a.s} \tilde{W}_\infty, \quad \text{and} \quad \frac{B_n}{n} \xrightarrow{a.s} \tilde{B}_\infty.$$

Example: In the original Pòlya urn model [2], when $m = 1$ and $X = C$ (deterministic), the random variable $\tilde{W}_\infty/C$ has a $Beta(\frac{B_0}{C}, \frac{W_0}{C})$ distribution [7, 20]. Whereas, M.R. Chen and M. Kuba [11] considered the case when $X = C$ (non random) and $m > 1$. They gave moments of all orders of W_n and proved that $\tilde{W}_\infty$ cannot be an ordinary $Beta$ distribution.

Remark: Suppose that the random variable X has moments of all orders, let $m_k = E(X^k)$, for $k \geq 1$. We have, almost surely, $W_n \leq T_n$ then, by Minskowski inequality, we obtain $\mathbb{E}(W_n^{2k}) \leq (mn)^{2k}\mathbb{E}(X^{2k})$. Using Carleman's condition we conclude that, if $\sum_{k\geq 1} \mu_{2k}^{-\frac{1}{2k}} = \infty$, then the random variable $\tilde{W}_\infty$ is determined by its moments. Unfortunately, till now we still unable to give exact expressions of moments of all orders of W_n. But, we can characterize the distribution of $\tilde{W}_\infty$ in the case when the variable X is bounded.

Lemma 4. *Assume that X is a bounded random variable, then, for fixed W_0, B_0 and m the random variable $\tilde{W}_\infty$ is absolutely continuous.*

The proof that $\tilde{W}_\infty$ is absolutely continuous is very close to that of Theorem 4.2 in [12]. We give the main proposition to make the proof clearer.

Proposition 3. *(see [12]) Let Ω_ℓ be a sequence of increasing events such that $\mathbb{P}(\cup_{\ell\geq 0}\Omega_\ell) = 1$. If there exists nonnegative Borel measurable function $\{f_\ell\}_{\ell\geq 1}$ such that $\mathbb{P}\Big(\Omega_\ell \cap \tilde{W}_\infty^{-1}(B)\Big) = \int_B f_\ell(x)dx$ for all Borel sets B, then, $f = \lim_{l\to+\infty} f_\ell$ exists almost everywhere and f is the density of $\tilde{W}_\infty$.*

Let $(\Omega, \mathcal{F}, \mathcal{P})$ be a probability space. Suppose that there exists a constant A such that, we have almost surely, $X \leq A$.

Lemma 5. *Define the events*

$$\Omega_\ell := \{W_\ell \geq mA \text{ and } B_\ell \geq mA\},$$

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

Next, we just need to show that the restriction of $\tilde{W}_\infty$ on $\Omega_{\ell,j} = \{\omega; W_\ell(\omega) = j\}$ has a density for each j, with $Am \leq j \leq T_{\ell-1}$. Let $(p_c)_{c\in \text{supp}(X)}$ the distribution of X.

Lemma 6. *For a fixed $\ell > 0$, there exists a positive constant κ, such that, for every $c \in$ supp(X), $n \geq \ell + 1$, $Am \leq j \leq T_{\ell-1}$ and $k \leq Am(n + 1)$, we have*

$$\sum_{i=0}^{m} \mathbb{P}(W_{n+1} = j + k | W_n = j + k - ci) \leq p_c(1 - \frac{1}{n} + \frac{\kappa}{n^2}). \tag{26}$$

Proof. According to Lemma 4.1 in [12], for $Am \leq j \leq T_{\ell-1}$, $n \geq \ell$ and $k \leq Am(n + 1)$, the following holds:

$$\sum_{i=0}^{m} \binom{j + c(k - i)}{i}\binom{T_n - j - c(k - i)}{m - i} = \frac{T_n^m}{m!} + \frac{(1 - m - 2c)T_n^{m-1}}{2(m - 1)!} + ..., \tag{27}$$

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 Equation (27) to our model we have

$$\begin{aligned}
u_{n,k}(c) &= p_c \sum_{i=0}^{m} \binom{j+k}{i}\binom{T_n - j - k}{m - i}\binom{T_n}{m}^{-1} \\
&= p_c \binom{T_n}{m}^{-1}\Big(\frac{T_n^m}{m!} + \frac{(1 - m - 2c)}{(m - 1)!}T_n^{m-1} + \dots\Big)\Big(\frac{T_n^m}{m!} + \frac{(1 - m)}{2(m - 1)!}T_n^{m-1} + \dots\Big)^{-1} \\
&\overset{a.s}{=} p_c\Big(1 - \frac{1}{n} + O\Big(\frac{1}{n^2}\Big)\Big).
\end{aligned} \tag{28}$$

□

Later, we will limit the proof by mentioning the main steps. For a fixed ℓ and $n \geq \ell + 1$, we denote by $v_{n,j} = \max\limits_{0\leq k\leq Amn} \mathbb{P}\big(W_{\ell+n} = j + k | W_\ell = j\big)$. We have the following inequality:

$$\begin{aligned}
v_{n+1,j} \quad &\leq \quad \max_{0\leq k\leq Am(n+1)} \Big\{\sum_{i=0}^{m}\sum_{c\in\text{supp}(X)} \mathbb{P}(W_{\ell+n+1} = j+k | W_{\ell+n} = j+k-ci)\Big\} \\
&\leq \quad \max_{0\leq k\leq Am(n+1)} \Big\{\sum_{i=0}^{m}\sum_{c\in\text{supp}(X)} \mathbb{P}(W_{\ell+n+1} = j+k | W_{\ell+n} = j+k-ci) \\
&\quad\quad \times \mathbb{P}(W_{\ell+n} = j+k-ci | W_\ell = j)\Big\} \\
&\leq \quad \max_{0\leq k\leq Am(n+1)} \sum_{i=0}^{m}\sum_{c\in\text{supp}(X)} \mathbb{P}(W_{\ell+n+1} = j+k | W_{\ell+n} = j+k-ci) \\
&\quad\quad \times \max_{0\leq \tilde{k}\leq Amn} \mathbb{P}\big(W_{\ell+n} = j+\tilde{k} | W_\ell = j\big) \\
&\leq \quad \sum_{c\in\text{supp}(X)} p_c\Big(1 - \frac{1}{n+l} + \frac{\kappa}{(n+l)^2}\Big)v_{n,j} \\
&\quad\quad = \Big(1 - \frac{1}{n+l} + \frac{\kappa}{(n+l)^2}\Big)v_{n,j}.
\end{aligned}$$

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

$$\max_{0\leq k\leq m(n-l)} \mathbb{P}\big(W_n = j + k | W_l = j\big) \leq \prod_{i=\ell}^{n}\Big(1 - \frac{1}{i} + \frac{\kappa}{i^2}\Big) \leq \frac{C(\ell)}{n}. \tag{29}$$

The rest of the proof follows.

□

Proof of Theorem 3. Consider the urn model evolving by the matrix $Q_n = \begin{pmatrix} 0 & X_n \\ Y_n & 0 \end{pmatrix}$. According to Equation (1), we have the following recursions:

$$W_{n+1} = W_n + X_{n+1}(m - \xi_{n+1}) \quad \text{and} \quad T_{n+1} = T_n + mX_{n+1} + \xi_{n+1}(Y_{n+1} - X_{n+1}). \tag{30}$$

Lemma 7. *The proportion of white balls after n draws, Z_n, satisfies the stochastic algorithm defined by (20), where $f(x) = m(\mu_X - \mu_Y)x^2 - 2\mu_X mx + \mu_X m$, $\gamma_n = \frac{1}{T_n}$ 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. We check the conditions of Definition 1, indeed,

(i) recall that $T_n = T_0 + m\sum_{i=1}^n X_i + \sum_{i=1}^n \xi_i(Y_i - X_i)$, then $\frac{T_n}{n} \leq \frac{T_0}{n} + \frac{m}{n}\sum_{i=1}^n X_i + \frac{m}{n}\sum_{i=1}^n |Y_i - X_i|$. By the strong law of large numbers we have $\frac{T_n}{n} \leq m(\mu_X + \mu_{|Y-X|}) + 1$. On the other hand, we have $T_n \geq \min\limits_{1\leq i\leq n}(X_i, Y_i)mn$, thus, the following bound holds

$$\frac{1}{(m(\mu_X + \mu_{|Y-X|}) + 1)n} \leq \frac{1}{T_n} \leq \frac{1}{m\min\limits_{1\leq i\leq n}(X_i, Y_i)n},$$

let $c_l = \frac{1}{(m(\mu_X + \mu_{|Y-X|}) + 1)}$ and $c_u = \frac{1}{m\min\limits_{1\leq i\leq n}(X_i, Y_i)}$,

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

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

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

□

Proposition 4. *The proportion of white balls in the urn after n draws, Z_n, satisfies as n tends to infinity*

$$Z_n \xrightarrow{a.s} z := \frac{\sqrt{\mu_X}}{\sqrt{\mu_X}+\sqrt{\mu_Y}}. \tag{31}$$

Proof. The proportion of white balls in the urn satisfies the stochastic approximation algorithm defined by Equation (20). As the function f is continuous, by Theorem 5, the process Z_n converges almost surely to $z = \frac{\sqrt{\mu_X}}{\sqrt{\mu_X}+\sqrt{\mu_Y}}$, the unique zero of f with negative derivative.

□

Next, we give an estimate of T_n, the total number of balls in the urn after n draws, in order to describe the asymptotic of the urn composition. By Equation (30), we have

$$\frac{T_n}{n} = \frac{T_0}{n} + \frac{m}{n}\sum_{i=1}^{n} X_i + \frac{m(\mu_Y-\mu_X)}{n}\sum_{i=1}^{n} Z_{i-1} + \frac{1}{n}\sum_{i=1}^{n}\Big[\xi_i(Y_i-X_i) - \mathbb{E}[\xi_i(Y_i-X_i)|\mathcal{F}_{i-1}]\Big].$$

Since $(X_i)_{i\geq 1}$ are iid 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 Cesáro lemma, we conclude that $\frac{1}{n}\sum_{i=1}^{n} Z_{i-1}$ converges almost surely, as n tends to infinity, to z. Finally, we prove that last term in the right side tends to zero, as n tends to infinity. In fact, let $G_n = \sum_{i=1}^{n}\Big[\xi_i(Y_i-X_i) - \mathbb{E}[\xi_i(Y_i-X_i)|\mathcal{F}_{i-1}]\Big]$, then $(G_n, \mathcal{F}_n)$ is a martingale difference sequence such that

$$\frac{<G>_n}{n} = \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}[\nabla G_i^2|\mathcal{F}_{i-1}],$$

where $\nabla G_n = G_n - G_{n-1} = \xi_n(Y_n - X_n) - \mathbb{E}[\xi_n(Y_n-X_n)|\mathcal{F}_{n-1}]$ and $<G>_n$ denotes the quadratic variation of the martingale.
By a simple computation, we have the almost sure convergence of $\mathbb{E}[\nabla G_i^2|\mathcal{F}_{i-1}]$ to $(mz(1-z)+m^2z^2)(\sigma_Y^2+\sigma_X^2)$. Therefore, Cesáro lemma ensures that, $\frac{<G>_n}{n}$ converges to $(mz(1-z)+m^2z^2)(\sigma_Y^2+\sigma_X^2)$, it follows that $\frac{G_n}{n} \xrightarrow{a.s} 0$. Thus, for n large enough we have

$$\frac{T_n}{n} \xrightarrow{a.s} m\sqrt{\mu_X}\sqrt{\mu_Y}. \tag{32}$$

In view of Equation (32), we describe the asymptotic behavior of the urn composition after n draws. One can write $\frac{W_n}{n} = \frac{W_n}{T_n}\frac{T_n}{n}$ and $\frac{B_n}{n} \stackrel{a.s}{=} \frac{B_n}{T_n}\frac{T_n}{n}$, using Equations (31, 32) and Slutsky theorem, we have, as n tends to infinity, $\frac{W_n}{n} \xrightarrow{a.s} m\sqrt{\mu_X}\sqrt{\mu_Y}z$ and $\frac{B_n}{n} \xrightarrow{a.s} m\sqrt{\mu_X}\sqrt{\mu_Y}(1-z)$.

Proof of claim 2

Later, we aim to apply Theorem 6. In our model, we have $\gamma_n = \frac{1}{T_n}$, then we need to control the following asymptotic behaviors

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

In fact, recall that $\frac{n}{T_n}$ converges almost surely to $\frac{1}{m\sqrt{\mu_X}\sqrt{\mu_Y}}$ and $\mathbb{E}[\Delta M_{n+1}^2|\mathcal{F}_n] = \mathbb{E}[D_{n+1}^2|\mathcal{F}_n] + \mathbb{E}[D_{n+1}|\mathcal{F}_n]^2$. Since $\mathbb{E}[D_{n+1}|\mathcal{F}_n]^2$ converges almost surely to $f(z)^2 = 0$, we have,

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

Using the fact that $\mathbb{E}[\xi_{n+1}^2|\mathcal{F}_n] = mZ_n(1-Z_n)\frac{T_n-m}{T_n-1} + m^2Z_n^2$ and that Z_n converges almost surely to z, we conclude that $\mathbb{E}[D_{n+1}^2|\mathcal{F}_n]$ converges almost surely to $G(z) > 0$. Applying Theorem 6, we obtain the following

$$\sqrt{n}(Z_n - z) \xrightarrow{\mathcal{L}} \mathcal{N}\Big(0, \frac{G(z)}{3m^2\mu_X\mu_Y}\Big). \tag{33}$$

But, we can write $\frac{W_n - zT_n}{\sqrt{n}} = \sqrt{n}\big(\frac{W_n}{T_n} - z\big)\frac{T_n}{n}$. Thus, it is enough to use Slutsky theorem to conclude the proof.

□

Proof of Theorem 4. Consider the urn model defined by Equation (1) with $Q_n = \begin{pmatrix} X_n & 0 \\ 0 & Y_n \end{pmatrix}$. The process of the urn satisfies the following recursions:

$$W_{n+1} = W_n + X_{n+1}\xi_{n+1} \quad \text{and} \quad T_{n+1} = T_n + mY_{n+1} + \xi_{n+1}(X_{n+1} - Y_{n+1}). \tag{34}$$

Lemma 8. *If $\mu_X \neq \mu_Y$, the proportion of white balls in the urn after n draws satisfies the stochastic algorithm defined by Equation (20) where $\gamma_n = \frac{1}{T_n}$, $f(x) = m(\mu_Y - \mu_X)x(x-1)$ 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) Recall that $T_n = T_0 + m\sum_{i=1}^n Y_i + \sum_{i=1}^n \xi_i(X_i - Y_i)$, then via the strong law of large numbers we have $|\frac{T_n}{n}| \leq m\mu_Y + m\mu_{|X-Y|} + 1$. On the other hand, we have $T_n \geq \min_{1\leq i\leq n}(X_i, Y_i)mn$, thus,

$$\frac{1}{(m\mu_Y + m\mu_{|X-Y|})n} \leq \frac{1}{T_n} \leq \frac{1}{\min\limits_{1\leq i\leq n}(X_i, Y_i)mn},$$

let $c_l = \frac{1}{(m\mu_Y + m\mu_{|X-Y|})}$ and $c_u = \frac{1}{\min\limits_{1\leq i\leq n}(X_i, Y_i)m}$,

(ii) $\mathbb{E}[\Delta M_{n+1}^2|\mathcal{F}_n] \leq (2m + m^2)(4\mu_{X^2} + \mu_{Y^2}) + 3m^2\mu_{Y^2} + 2m^2\mu_X + 2m^2\mu_X\mu_Y + 4m^2(\mu_X - \mu_Y)^2 = K_u$,

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

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

□

Proposition 5. *The proportion of white balls in the urn after n draws, Z_n, satisfies almost surely*

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

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

Proof of Proposition 5. Recall that, if $\mu_X \neq \mu_Y$, Z_n satisfies the stochastic algorithm defined in Lemma 8. As the function f is continuous, by Theorem 6 we conclude that Z_n converges almost surely 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 $P_{n+1} = X_{n+1}\xi_{n+1} - Z_n\big(mY_{n+1} + \xi_{n+1}(X_{n+1} - Y_{n+1})\big)$. Since $\mathbb{E}[P_{n+1}|\mathcal{F}_n] = 0$, then Z_n is a positive martingale which converges almost surely to a positive random variable $\tilde{Z}_\infty$. As a consequence, we have

Corollary 2. *The total number of balls in the urn, T_n, satisfies as n tends to infinity*

if $\mu_X \geq \mu_Y$

$$\frac{T_n}{n} \xrightarrow{a.s} m\mu_X.$$

Proof. In fact, let $M_n = \sum_{i=1}^n \xi_i(X_i - Y_i) - \mathbb{E}[\xi_i(X_i - Y_i)|\mathcal{F}_{i-1}]$, we have

$$\begin{aligned}\frac{T_n}{n} &= \frac{T_0}{n} + \frac{m}{n}\sum_{i=1}^n Y_i + \frac{1}{n}\sum_{i=1}^n \xi_i(X_i - Y_i)\\ &= \frac{T_0}{n} + \frac{m}{n}\sum_{i=1}^n Y_i + \frac{m(\mu_X - \mu_Y)}{n}\sum_{i=1}^n Z_{i-1} + \frac{M_n}{n}.\end{aligned}$$

As it was proved in claim 1 of Theorem 3, we show that, as n tends to infinity, we have $\frac{M_n}{n} \xrightarrow{a.s} 0$. Recall that, if $\mu_X > \mu_X$, Z_n converges almost surely to 1. 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 $\mu_Y = \mu_X$. □

Using the results above, the convergence of the normalized number of white balls follows immediately. Indeed, if $\mu_X > \mu_Y$, we have, as n tends to infinity,

$$\frac{W_n}{n} = \frac{W_n}{T_n}\frac{T_n}{n} \xrightarrow{a.s} m\mu_X,$$

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 $\prod_{i=1}^{n-1}(1+\frac{m\mu_Y}{T_i}) \simeq An^\rho$; $\rho = \frac{\mu_Y}{\mu_X}$. Then, as n tends to infinity we have

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

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

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

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

□

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

$$T_n \overset{a.s}{=} m\mu_Y n + o(n), \quad W_n = \tilde{W}_\infty n^\sigma + o(n) \quad \text{and} \quad B_n = m\mu_Y n + o(n),$$

where $\tilde{W}_\infty$ is a positive random variable and $\sigma = \frac{\mu_X}{\mu_Y}$.

References

[1] C. Mailler, N. Lasmer and O. Selmi.(2018). Multiple drawing multi-color urns by stochastic approximation.*Journal of Applied Probability.* 55(1):254-281.

[2] F. Eggenberger and G. Pólya. (1923). Über die statistik verkeletter vorge. *Zeitschift für Angewandte Mathematik und Mechanic.* 1:279-289.

[3] G. Pagés and S. Laruelle. (2015). Randomized urns models revisited using stochastic approximation. *Annals of Applied Probability.* 23(4): 1409-1436.

[4] H. Renlund. (2010). Generalized Polya urns via stochastic approximation. arxiv: 1002.3716v1.

[5] H. Renlund. (2011). Limit theorem for Stochastic approxiamtion algorithm. arxiv :1102.4741v1.

[6] H. Mahmoud. (2004). Random spouts as internet model and Pólya processes. *Actainformatica,* 41: 1-18.

[7] K.B. Athreya and P.E. Ney. (1972). Branching Processes. *Springer-Verlag, Berlin.*

[8] L.J. Wei. (1978). An application of an urn model to the design of sequential controlled clinical trials. *Journal of American Statistics Association.* 73:559-563.

[9] M. Duflo.(1997). Random iterative models. *Springer-Verlag, Berlin.*

[10] M. Kuba, H. Mahmoud and A. Panholzer. (2013). Analysis of a generalized Friedman's urn with multiple drawings. *Discrete Applied Mathematics.* 161:2968-2984 .

[11] M. R Chen and M. Kuba. (2013). On generalized Polya urn models. *Theory of probability and its application.* 40: 1169-1186.

[12] M. R Chen and C. Z Wei. (2005). A new urn model. *Applied Probability.* 4(42): 964-976.

[13] M. Kuba and H. Sulzbach. (2016). On martingale tail sums in affine two-color urn models with multiple drawings. arXiv:1509.09053.

[14] M. Kuba and H. Mahmoud. (2016). Two-colour balanced affine urn models with multiple drawings I: Central limit theorem. arXiv :1503.09069.

[15] M. Kuba and H. Mahmoud. (2016). Two-colour balanced affine urn models with multiple drawings II: large-index and triangular urns. arXiv: 1509.09053.

[16] N.L. Johnson, S. Kotz. (1977). Urn models and their application. *John Wiley Son.*

[17] R. Aguech.(2009). Limit Theorems for Random Triangular Urns Schemes. *Journal of Applied Probability.* 46(3):827-843.

[18] R. Aguech, N. Lasmer, O. Selmi.(2018). A generalized urn model with multiple drawing and random addition. *annals of the instituteof Statistical Mathematics.* 1-20.

[19] R.N. Goldman.(1985). Polya's urn model and computer aided geometry design.*SIAM Journal on Algebraic Discrete Methods,* 6(1), 1-28.

[20] S. Janson.(2006).Limit theorems of triangular urn schemes.*Probability Theory and Related Fields,* 134(3), 417-452.

[21] Z. Lin and C. Lu.(1996) Limit Theory for Mixing Dependent Random Variables. *Kluwer Academic Publishers, Boston.*