Ann Inst Stat Math (2019) 71:389–408
https://doi.org/10.1007/s10463-018-0651-3

A generalized urn with multiple drawing and random addition

Aguech Rafik[1,2] · Lasmar Nabil[3] · Selmi Olfa[2]

Received: 25 September 2016 / Revised: 22 October 2017 / Published online: 2 March 2018

Abstract In this paper, we consider an unbalanced urn model with multiple drawing. At each discrete time step n, we draw m balls at random from an urn containing white and blue balls. The replacement of the balls follows either opposite or self-reinforcement rule. Under the opposite reinforcement rule, we use the stochastic approximation algorithm to obtain a strong law of large numbers and a central limit theorem for W_n: the number of white balls after n draws. Under the self-reinforcement rule, we prove that, after suitable normalization, the number of white balls W_n converges almost surely to a random variable W_∞ which has an absolutely continuous distribution.

Keywords Unbalanced urn · Stochastic approximation · Martingale · Maximal inequality

✉ Aguech Rafik
raguech@ksu.edu.sa; rafik.aguech@ipeit.rnu.tn

Lasmar Nabil
nabillasmar@yahoo.fr

Selmi Olfa
selmiolfa3@yahoo.fr

1 Department of Statistics and Operation Research, King Saoud University, Riyadh, Riyadh 11692, Kingdom of Saudi Arabia

2 Département de Mathématiques, Faculté des Sciences de Monastir, Avenue de l’environnement, 5019 Monastir, Tunisia

3 Département de Mathématiques, Institut Préparatoire aux Études d’ingénieurs de Monastir, Avenue de l’environnement, 5019 Monastir, Tunisia

1 Introduction

The classic urn model was introduced by Eggenberger and Polya (1923). This urn contains balls of two different colors. A ball is drawn randomly, and it is returned to the urn with α balls of the same color. Kuba and Mahmoud (2016a), Chen and Kuba (2013), Chen and Wei (2005), Konzem and Mahmoud (2016), Tsukiji and Mahmoud (2001) study a generalized urn model. This model evolves by drawing randomly at each step n a sample of m balls ($m \in \{2, 3, \ldots\}$). Balls are added according to some prescribed rules depending on the colors of the drawn sample. Chen and Wei (2005) considered the *self-reinforcement* urn model where each ball in the sample is returned to the urn with a balls of the same color (a is an integer). In such model, the normalized number of white balls after n draws W_n converges, almost surely, to a nondegenerate random variable W_∞ with an absolutely continuous distribution. In the *opposite reinforcement* model each ball in the drawn sample is returned to the urn with a balls of the opposite color. Using different methods, such as martingales, moments and recurrences. Kuba et al. (2013) described the asymptotic normality of the normalized number of white balls. In a recent works, Kuba and Mahmoud (2016a, b), studied a new class of generalized urn model, which is the affine one: the number of white balls satisfies an affine recurrence. Using some stochastic algorithms results of Duflo (1997), Renlund (2010) obtains some asymptotic results for an unbalanced urn with $m = 1$. Stochastic algorithms methods have been also used by Pagés and Laruelle (2015) for a balanced urn when $m \geqslant 2$.

The dynamics of an urn process can simply describe numerous applications. They treated biology, computer data, physics, finance, etc. Therefore, it was the interest of many authors Baggchi and Paul (1985), Chauvin et al. (2011), Flajolet et al. (2005) Athreya and Ney (1972) who treated balanced urn model where a fixed number of balls is added at each step add at each step.

In the present paper, we deal with an unbalanced urn model under both opposite- and self-reinforcement rules. The model is defined as follows: initially the urn contains W_0 white balls and B_0 blue balls such that $T_0 := W_0 + B_0 \geqslant m$. A sample of m balls is taken at random from the urn, and each white (resp blue) ball from the sample is returned to the urn with a (resp b) balls with either the same or the opposite color depending on the rule. Let W_n, B_n and T_n be, respectively, the number of white, blue and the total number of balls in the urn after n draws. Let $(\mathcal{F}_n)_{n\geqslant 0}$ be the $\sigma-$field generated by the first n draws and denote by ξ_n the number of white balls in the sample in the n^{th} draw. We have

$$P(\xi_n = j|\mathcal{F}_{n-1}) = \frac{\binom{W_{n-1}}{j}\binom{B_{n-1}}{m-j}}{\binom{T_{n-1}}{m}}, \quad (0 \leqslant j \leqslant m). \tag{1}$$

we characterize the evolution of the urn by the following stochastic recursion:

$$\begin{pmatrix} W_{n+1} \\ B_{n+1} \end{pmatrix} \overset{\mathcal{D}}{=} \begin{pmatrix} W_n \\ B_n \end{pmatrix} + Q \begin{pmatrix} \xi_{n+1} \\ m - \xi_{n+1} \end{pmatrix}, \tag{2}$$

where for the self-reinforcement rule Q is given by $Q := \begin{pmatrix} a & 0 \\ 0 & b \end{pmatrix}$, and for the opposite reinforcement rule the expression of Q is $\begin{pmatrix} 0 & a \\ b & 0 \end{pmatrix}$. The total number of balls is given by

$$T_{n+1} = T_n + \left\langle Q \begin{pmatrix} \xi_{n+1} \\ m - \xi_{n+1} \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \end{pmatrix} \right\rangle, \quad \forall n \geqslant 0, \tag{3}$$

the notation $\langle , \rangle$ is for the canonical inner product in $\mathbb{R}^2$. Section 2 is devoted to the opposite reinforcement model. We give the moments, the strong law of large number as well as the asymptotic normality of $\frac{W_n}{T_n}$, which is the proportion of white balls. In Sect. 3, we treat the self-reinforcement model as we give some properties of the limit law of the normalized number of white balls.

2 Opposite reinforcement model

In this section, we consider a two-color urn under the opposite reinforcement rule. We show that the proportion of white balls satisfies a stochastic algorithm (We refer the reader to the monographs Duflo (1997) or Benveniste et al. (1990) for a thorough introduction and a comprehensive overview of this discipline)) that, under some conditions, converges almost surely to a set of stable points of a function h.

Theorem 1 (Renlund 2010) *If a given sequence $(X_n)_{n\geqslant 0}$ satisfies*

$$X_{n+1} - X_n = \gamma_{n+1}(f(X_n) + U_{n+1}), \tag{4}$$

where $(\gamma_n)_{n\geqslant 1}$ and $(U_n)_{n\geqslant 1}$ are two $\mathcal{F}_n$-measurable sequences of random variables and f is a continuous function from $[0, 1]$ onto $\mathbb{R}$ such that $f \not\equiv 0$. Assume that, almost surely,

$$\frac{c_1}{n} \leqslant \gamma_n \leqslant \frac{c_2}{n}, \qquad |U_n| \leqslant K_u \qquad |f(X_n)| \leqslant K_f, \quad \text{and} \quad \mathbb{E}[\gamma_{n+1}U_{n+1}|\mathcal{F}_n] \leqslant K_e\gamma_n^2,$$

where the constants c_1, c_2, K_u, K_f, and $K_e$are positive real numbers. Then, we have $X_\infty := \lim_{n\to+\infty} X_n$ exists almost surely and $f(X_\infty) = 0$.

Definition 1 A zero θ of a differentiable function h is called a stable zero, if and only if, all eigenvalues of $Dh(\theta)$ are negative.

Lemma 1 *Let $Z_n = \frac{W_n}{T_n}$ be the proportion of white balls after n draws. The sequence $(Z_n)_{n\geqslant 0}$ satisfies the following stochastic algorithm:*

$$Z_{n+1} - Z_n = \gamma_{n+1}\Big(h(Z_n) + \Delta M_{n+1}\Big), \tag{5}$$

where

$$h(x) = m(a-b)x^2 - 2amx + am, \quad \Delta M_{n+1} = Y_{n+1} - \mathbb{E}(Y_{n+1}|\mathcal{F}_n),$$

with $Y_{n+1} = am + (a-b)\xi_{n+1}Z_n - a\xi_{n+1} - amZ_n$.

Proof Recall that

$$W_{n+1} = W_n + a(m - \xi_{n+1}), \quad T_{n+1} = T_n + am + (b-a)\xi_{n+1},$$

and we have

$$\begin{aligned} Z_{n+1} - Z_n &= \frac{1}{T_{n+1}}\big(W_n + a(m-\xi_{n+1}) - Z_n(T_n + am + (b-a)\xi_{n+1})\big) \\ &= \frac{1}{T_{n+1}}\big(am - a\xi_{n+1} - Z_n(am + (b-a)\xi_{n+1})\big) \\ &= \frac{Y_{n+1}}{T_{n+1}}. \end{aligned}$$

Note that for either the model with or without replacement we have $\mathbb{E}(\xi_{n+1}|\mathcal{F}_n) = mZ_n$, which implies that

$$\mathbb{E}(Y_{n+1}|\mathcal{F}_n) = m(a-b)Z_n^2 - 2amZ_n + am,$$

and we conclude the result. □

Proposition 1 *The proportion of white balls Z_n in an unbalanced urn model under the opposite reinforcement rule converges almost surely to x_1, where $x_1 = \dfrac{\sqrt{a}}{\sqrt{a}+\sqrt{b}}$.*

Proof We now aim to apply Theorem 1. Recall that the process Z_n satisfies (4) with $\gamma_n = \frac{1}{T_n}$, $f \equiv h$ and $U_n = \Delta M_{n+1}$. We have $T_0 + mn\min(a,b) \leqslant T_n \leqslant T_0 + mn\max(a,b)$. Then γ_n satisfies the following bound

$$\frac{c_1}{n} \leqslant \frac{1}{T_n} \leqslant \frac{c_2}{n},$$

where $c_1 = \frac{1}{1+m\max(a,b)}$ and $c_2 = \frac{1}{m\min(a,b)}$, $\forall n \geqslant [T_0]$. Since the process Z_n is bounded by one we have

$$|h(Z_n)| \leqslant 3am + m|a-b|,$$

and

$$|\Delta M_n| \leqslant 6am + 2m|a-b|.$$

On the other hand

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

Indeed $h(0) = am > 0$ and $h(1) = -mb < 0$, according to Theorem 1 the process Z_n converges almost surely to a stable zero of h. Note that h has two zeros $x_1 = \frac{\sqrt{a}}{\sqrt{a}+\sqrt{b}}$ and $x_2 = \frac{\sqrt{a}}{\sqrt{a}-\sqrt{b}}$ with $h'(x_1) < 0$ and $h'(x_2) > 0$, hence x_1 is the stable one. □

Corollary 1 *The normalized total number of balls in the urn after n draws satisfies when, $nrightarrow + \infty$,*

$$\frac{T_n}{n} \xrightarrow{a.s} \sqrt{abm}. \tag{6}$$

Proof Recall that the total number of balls satisfies

$$\frac{T_n}{n} = \frac{T_0}{n} + am + \frac{(b-a)}{n}\sum_{k=1}^{n}\left(\xi_k - m\frac{W_{k-1}}{T_{k-1}}\right) + \frac{m(b-a)}{n}\sum_{k=1}^{n}\frac{W_{k-1}}{T_{k-1}}. \tag{7}$$

Let us denote by $G_n = \sum_{k=1}^{n}(\xi_k - m\frac{W_{k-1}}{T_{k-1}})$, then $(G_n, \mathcal{F}_n)_{n\geqslant 0}$ is a martingale with martingale difference $\nabla G_n = G_n - G_{n-1} = \xi_n - m\frac{W_{n-1}}{T_{n-1}}$. Note that the quadratic variation process is defined as follow

$$\begin{aligned} <G>_n &= \sum_{k=1}^{n}\mathbb{E}(\nabla G_k^2|\mathcal{F}_{k-1}) \\ &= \sum_{k=1}^{n}\mathbb{V}ar(\xi_k|\mathcal{F}_{k-1}) \\ &= \sum_{k=1}^{n} m\frac{W_{k-1}}{T_{k-1}}\frac{B_{k-1}}{T_{k-1}}\frac{T_{k-1}-m}{T_{k-1}-1}. \end{aligned}$$

Since $m\frac{W_n}{T_n}\frac{B_n}{T_n}\frac{T_n-m}{T_n-1}$ converges almost surely to $mx_1(1-x_1)$, by Cesaro's lemma, $\frac{<G>_n}{n}$ converges almost surely to the same limit,and, we get $\frac{G_n}{n} \to 0$ almost surely when n tends to infinity. Therefore, by Eq. (7), $\frac{T_n}{n}$ converges almost surely to $m\sqrt{ab}$. □

In the sequel, we shall investigate the rate of convergence of T_n.

Proposition 2 *The total number of balls in the urn satisfies:*

$$T_n = \sqrt{abm}n + o(n^{\frac{1}{2}+\delta}) \quad \textit{almost surely}, \tag{8}$$

with δ being any arbitrarily number in $]0, \frac{1}{4}[$.

Proof In view of the relation (2), the variance of W_n satisfies the following recurrence:

$$\begin{aligned}\mathbb{V}ar(W_{n+1}) &= \mathbb{V}ar(W_n) + a^2\,\mathbb{V}ar\,(\xi_{n+1}) - 2am\mathbb{C}ov\left(\frac{W_n}{T_n}, W_n\right)\\ &= \left(1 - \frac{2}{n}\sqrt{\frac{a}{b}}(1+o(1))\right)\mathbb{V}ar(W_n) + O(1)\\ &= \alpha_n\,\mathbb{V}ar(W_n) + \beta_n,\end{aligned}$$

where $\alpha_n = \left(1 - \frac{2}{n}\sqrt{\frac{a}{b}}(1+o(1))\right)$ and $\beta_n = O(1)$. We obtain

$$\mathbb{V}ar(W_n) = \left(\prod_{k=1}^{n-1}\alpha_k\right)\left(\mathbb{V}ar(W_1) + \sum_{k=1}^{n-1}\frac{\beta_k}{\prod_{j=1}^{k-1}\alpha_j}\right),$$

and there exists some positive constant C, such that $\prod_{k=1}^{n-1}\alpha_k = \frac{C}{n^{2\sqrt{\frac{a}{b}}+o(1)}}$. Then $\mathbb{V}ar(W_n) = O(n^{1+\delta})$, for any $\delta > 0$. Furthermore we have

$$aB_n + bW_n = abmn + aB_0 + bW_0, \tag{9}$$

subsequently,

$$T_n = \left(1 - \frac{b}{a}\right)W_n + bmn + \left(B_0 + \frac{b}{a}W_0\right). \tag{10}$$

We conclude that $\mathbb{V}ar(T_n) = \left(1 - \frac{b}{a}\right)^2\mathbb{V}ar(W_n) = O(n^{1+\delta})$. Thus, by Longnecker and Serfling (1977), there exists some positive constant A such that

$$\mathbb{E}\left(\left(\max_{1\leqslant k\leqslant n}(T_k - \mathbb{E}(T_k))\right)^2\right) \leqslant A\sum_{k=1}^{n}k^{\delta}. \tag{11}$$

For $a_n = n^{\delta}$ and $b_n = n^{\frac{1}{2}+\delta}$ the series $\sum_{n\geqslant 1}\frac{a_n}{b_n^2}$ is convergent, and by Eq. (11) we can apply the generalized form of the strong law of large number in Fazekas and Klesov (2001) and we get

$$T_n = \mathbb{E}(T_n) + o\left(n^{\frac{1}{2}+\delta}\right) \text{ almost surely.} \tag{12}$$

By relation (9), the expectation of T_n satisfies

$$\mathbb{E}(T_{n+1}) = \mathbb{E}(T_n) + abm^2\frac{n}{\mathbb{E}(T_n)} + o\left(n^{\delta-\frac{1}{2}}\right). \tag{13}$$

Let $\gamma = m^2ab$ and $u_n = \dfrac{\mathbb{E}(T_n)}{\gamma n} - 1$. The sequence $(u_n)_{n\geqslant 0}$ satisfies the following recurrence:

$$n^2\gamma^2(u_n+1)^2 = (n-1)^2\gamma^2(u_{n-1}+1)^2 + 2\gamma(n-1) + o(n^{\frac{1}{2}+\delta}),$$

which leads to $(u_n+1)^2 = \dfrac{1}{\gamma} + o(n^{\delta-\frac{1}{2}})$, and the result follows. □

We use the result of (8) to give the asymptotic expansions of the moments of W_n.

Proposition 3 *The mean and the variance of the number of white balls in the urn after n draws satisfy*

$$\mathbb{E}(W_n) = m\sqrt{ab}x_1 n + o\left(n^{\delta+\frac{1}{2}}\right) \quad \textit{and} \quad \mathbb{V}ar(W_n)$$
$$= \frac{a^2 m\, x_1(1-x_1)^2}{1+x_1} n + o\left(n^{\delta+\frac{1}{2}}\right),$$

where δ *is arbitrarily in* $]0, \frac{1}{4}[$.

Proof Using Eqs. (10, 12) and proposition 8, we obtain $\mathbb{E}(W_n)$. On the other hand, recall that W_n satisfies the recurrence (2), then we have

$$\mathbb{V}ar(\xi_{n+1}) = \mathbb{E}\left(m\frac{W_n}{T_n}\left(1-\frac{W_n}{T_n}\right)\frac{T_n-m}{T_n-1}\right) + m^2\,\mathbb{V}ar\left(\frac{W_n}{T_n}\right)$$
$$= mx_1 - mx_1^2 + \frac{\mathbb{V}ar(W_n)}{abn^2} + o\left(n^{\delta-\frac{1}{2}}\right),$$

and

$$\mathrm{Cov}\left(W_n, \frac{W_n}{T_n}\right) = \frac{1}{m\sqrt{ab}n}\left(1+o\left(n^{\delta-\frac{1}{2}}\right)\right)\mathbb{V}ar(W_n).$$

We then obtain a recurrence for $\mathbb{V}ar(W_n)$:

$$\mathbb{V}ar(W_{n+1}) = \left(1-2\sqrt{\frac{a}{b}}\frac{1}{n} + o(n^{\delta-\frac{3}{2}})\right)\mathbb{V}ar(W_n) + a^2 m x_1(1-x_1) + o\left(n^{\delta-\frac{1}{2}}\right).$$

It follows that

$$\mathbb{V}ar(W_n) = \left(\prod_{k=1}^{n}\alpha'_k\right)\left(\mathbb{V}ar(W_0) + \sum_{k=0}^{n-1}\frac{\beta'_k}{\prod_{j=0}^{k}\alpha'_j}\right),$$

where $\alpha'_n = 1-2\sqrt{\frac{a}{b}}\frac{1}{n} + o(n^{\delta-\frac{3}{2}})$ and $\beta'_n = a^2 m x_1(1-x_1) + o(n^{\delta-\frac{1}{2}})$. There exists a positive constant α' such that $\prod_{k=1}^{n-1}\alpha'_k = \alpha'\frac{1}{n^{2\sqrt{\frac{a}{b}}}}(1+o(n^{\delta-\frac{1}{2}}))$. Hence

$$\sum_{k=0}^{n-1}\frac{\beta'_k}{\prod_{j=0}^{k-1}\alpha'_j} = \frac{a^2 m x_1(1-x_1)}{\alpha'\left(1+2\sqrt{\frac{a}{b}}\right)} n^{2\sqrt{\frac{a}{b}}+1}\left(1+o\left(n^{\delta-\frac{1}{2}}\right)\right)$$
$$= \frac{a^2 m x_1(1-x_1)^2}{\alpha'(1+x_1)} n^{2\sqrt{\frac{a}{b}}+1}\left(1+o\left(n^{\delta-\frac{1}{2}}\right)\right).$$ □

Corollary 2 *The number of white balls W_n in the urn after n draws satisfies, almost surely:*

$$W_n = m\sqrt{ab}\, x_1 n + o(\sqrt{n}\ln(n)). \tag{14}$$

Remark 1 The previous corollary allows us to give a better estimate of the mean as well as the variance of W_n involving the rate $o(\sqrt{n}\ln(n))$.

$$\mathbb{E}(W_n) = m\sqrt{ab}\, x_1 n + o(\sqrt{n}\ln(n)),$$
$$\mathbb{V}ar(W_n) = \frac{a^2 m\, x_1(1-x_1)^2}{1+x_1} n + o(\sqrt{n}\ln(n)).$$

2.1 Central limit theorem

Our aim in this subsection is to apply Renlund (2011) central limit theorem for stochastic algorithms. The result is expressed as follows:

Theorem 2 (Renlund 2011) *Let $(X_n)_{n\geqslant 0}$ be as in (4) with almost sure limit X_∞. Let $\hat{\gamma}_n := -n\gamma_n \frac{h(X_{n-1})}{X_{n-1}-X_\infty}$. Assume that $\hat{\gamma}_n$ converges almost surely to some limit $\hat{\gamma} > \frac{1}{2}$ and $\mathbb{E}[(n\gamma_n U_n)^2|\mathcal{F}_{n-1}] \to \sigma^2 > 0$, then we have*

$$\sqrt{n}(X_n - X_\infty) \xrightarrow{\mathcal{D}} \mathcal{N}\left(0, \frac{\sigma^2}{2\hat{\gamma}-1}\right).$$

Theorem 3 *Let Z_n be the proportion of white balls in the urn submitted to the opposite reinforcement rule after n draws. Then, the following holds:*

$$\sqrt{n}(Z_n - x_1) \xrightarrow{\mathcal{D}} \mathcal{N}\left(0, \frac{\sqrt{ab}}{3m(\sqrt{a}+\sqrt{b})^2}\right). \tag{15}$$

Proof Recall that Z_n satisfies (4) and it converges almost surely to x_1. In our model, we have $\frac{n}{T_n} \overset{a.s}{\to} \frac{1}{m\sqrt{ab}}$ and $\frac{-h(Z_n)}{Z_n - x_1} \overset{a.s}{\to} -h'(x_1) = 2m\sqrt{ab}$. By applying Theorem 2 with $\hat{\gamma}_n = -\frac{n}{T_n}\frac{h(Z_n)}{Z_n - x_1}$, we get $\hat{\gamma}_n \to \hat{\gamma} = 2 > \frac{1}{2}$. Now, we have

$$\begin{aligned}
&\mathbb{E}\Big(\Big(\frac{n}{T_n}\Delta M_n\Big)^2|\mathcal{F}_{n-1}\Big)\\
&= \mathbb{E}\Big(\Big(\frac{n}{T_n}\Big)^2((a-b)Z_{n-1}-a)^2(\xi_n - \mathbb{E}(\xi_n|\mathcal{F}_{n-1}))^2|\mathcal{F}_n\Big)\\
&= ((a-b)Z_{n-1}-a)^2\mathbb{E}\Big(\Big(\frac{n}{T_n}\Big)^2(\xi_n - \mathbb{E}(\xi_n|\mathcal{F}_{n-1}))^2|\mathcal{F}_n\Big)\\
&\overset{n\to+\infty}{\longrightarrow} \frac{\sqrt{ab}}{m(\sqrt{a}+\sqrt{b})^2} = \sigma^2.
\end{aligned}$$

□

Remark 2 Since the urn is not balanced, the total number of balls in the urn is random. The stochastic algorithms give us only strong convergence results. for the proportion of white balls. Unfortunately, we are unable to establish a central limit theorem for the number of white balls.

In the following paragraph, we will prove that the dependence between the variables ξ_n is weak. This result will indeed be used to describe the limit law of the number of white balls.

Lemma 2 *Let $q_{n_1} < \cdots < q_{n_r} < n$ be a sequence such that $q_{n_1} \to \infty$ and let $\hat{W}_n$ (resp $\hat{T}_n$) be the number of white balls (resp the total number of balls) after n draws conditioning on the event $\{\xi_{q_{n_1}} = m_{n_1}, \ldots, \xi_{q_{n_r}} = m_{n_r}\}$. Then*

$$\frac{\hat{W}_n}{\hat{T}_n} = x_1 + o\left(\frac{\ln(n)}{\sqrt{n}}\right), \quad (\textit{almost surely}) \tag{16}$$

Definition 2 Let $(\Omega, \mathcal{F}, P)$ be a probability space and $\mathcal{G}_1$ and $\mathcal{G}_2$ are sub σ-fields of $\mathcal{F}$. The strong mixing coefficient between two subalgebra $\mathcal{G}_1$ and $\mathcal{G}_2$ is given by

$$\alpha(\mathcal{G}_1, \mathcal{G}_2) = \sup\{|\mathbb{C}ov(u, v))|, 0 \leqslant u, v \leqslant 1\ \sigma(u) \subset \mathcal{G}_1,\ \sigma(v) \subset \mathcal{G}_2\}. \tag{17}$$

A sequence of random variables $(X_n)_n \geqslant 0$ on $(\Omega, \mathcal{F}, P)$ is strongly mixing if the sequence $(\alpha(n))_{n\geqslant 0}$ defined by

$$\alpha(n) = \sup_{k\geqslant 0}\{\alpha(\sigma(X_j, 0 \leqslant j \leqslant k), \sigma(X_j, j \geqslant k+n))\} \tag{18}$$

converges to 0.

Lemma 3 *The strong mixing coefficient of $(\xi_n)_{n\geqslant 1}$ satisfies $\alpha(n) = o\Big(\frac{\ln(n)}{\sqrt{n}}\Big)$.*

Proof Let U and V be two functions and $k \in \mathbb{N}$ such that $\sigma(U) \subset \sigma(Y_j; j \leqslant k)$ and $\sigma(V) \subset \sigma(Y_l; l \geqslant k+n)$, then U and V are written as a linear combination of simple functions of $\sigma(Y_j; j \leqslant k)$ and $\sigma(Y_l; l \geqslant k+n)$. Thus there exists $(a_h)_{1\leqslant h\leqslant r}$ and $(b_l)_{1\leqslant l\leqslant s}$ such that

$$U = \sum_{h=1}^{r} a_h \chi(\xi_i = m_i,\ i \in I_h) \quad \text{and} \quad V = \sum_{l=1}^{s} b_l \chi(\xi_j = m_j, j \in J_l), \tag{19}$$

where for $k \in \mathbb{N}$, $I_1, \ldots, I_r$ are subsets of $\{1, \ldots, k\}$, $J_1, \ldots, J_s$ are finite subsets of $\mathbb{N}\backslash\{1, \ldots, n+k\}$ and $m_i \in \{0, \ldots, m\}$. Note that for a fixed $s \in \{1, \ldots, r\}$ (resp a fixed $l \in \{1, \ldots, s\}$) we have $I_h = \{p_1, \ldots, p_{c_h}\}$ (resp $J_l = \{q_1, \ldots, q_{c_l}\}$). Let $w_{h,i} = \mathbb{P}\big(\xi_{q_i} = m_{q_i} | \xi_{q_j} = m_{q_j}; j \in I_h \bigcup\{1, \ldots, i-1\}\big)$ and $w_{1,i} = \mathbb{P}\big(\xi_{q_i} = m_{q_i} | \xi_{q_j} = m_{q_j}; 1 \leqslant j \leqslant i-1\big)$. We have

$$|\mathbb{C}ov(U, V)| \leqslant \sum_{h,l} |a_h b_l| \mathbb{P}\Big(\xi_i = m_i, i \in I_h\Big) \Big|\prod_{i=2}^{c_l} w_{h,i} - \prod_{l=2}^{c_l} w_{1,i}\Big|. \tag{20}$$

$$\leqslant \sum_{h,l} \sum_{i=2}^{c_l} |a_i b_j||w_{h,i} - w_{1,i}|. \tag{21}$$

The last inequality is due to the fact that $|\prod_{k=1}^n u_k - \prod_{k=1}^n v_k| \leqslant \sum_{k=1}^n |u_k - v_k|$,$\forall u_1, \ldots, u_n, v_1, \ldots, v_n$ in the unit circle. We have

$$\begin{aligned} w_{1,i} &= E\left(\binom{W_{q_i-1}}{m_{q_i}}\binom{B_{q_i-1}}{m-m_{q_i}}\binom{T_{q_i-1}}{m}^{-1} |\xi_{q_j} = m_{q_j}, 1 \leqslant j \leqslant i-1\right) \\ &= \binom{m}{m_{q_i}} E\left(\frac{(W_{q_i-1})_{m_{q_i}} (B_{q_i-1})_{m-m_{q_i}}}{(T_{q_i})_m} |\xi_{q_j} = m_{q_j}, 1 \leqslant j \leqslant i-1\right), \end{aligned}$$

where $(x)_n = x(x-1)\ldots(x-n+1) = x^n - \binom{n}{2}x^{n-1} + P_{n-2}(x)$ and $P_{n-2}(x)$ is a polynomial in x with degree $(n-2)$ such that $P_{n-2}(x) = 0$ for $n < 2$. For $\Gamma_{m_{q_i}}(q_i) = \frac{(W_{q_i-1})_{m_{q_i}} (B_{q_i-1})_{m-m_{q_i}}}{(T_{q_i-1})_m}$, we have

$$\begin{aligned} \Gamma_{m_{q_i}}(q_i) &= \frac{\left(W_{q_i-1}^{m_{q_i}} - \binom{m_{q_i}}{2} W_{q_i-1}^{m_{q_i}-1} + P_{m_{q_i}-2}(W_{q_i-1})\right)\left(B_{q_i-1}^{m-m_{q_i}} - \binom{m-m_{q_i}}{2} B_{q_i-1}^{m-m_{q_i}-1} + P_{m-m_{q_i}-2}(B_{q_i-1})\right)}{T_{q_i-1}^m - \binom{m}{2} T_{q_i-1}^{m-1} + P_{m-2}(T_{q_i-1})} \\ &\overset{a.s}{=} \Big(\Big(\frac{W_{q_i-1}}{T_{q_i-1}}\Big)^{m_{q_i}} \Big(\frac{B_{q_i-1}}{T_{q_i-1}}\Big)^{m-q_i} - \frac{\binom{m_{q_i}}{2}}{T_{q_i-1}} \Big(\frac{W_{q_i-1}}{T_{q_i-1}}\Big)^{m_{q_i}-1} \Big(\frac{B_{q_i-1}}{T_{q_i-1}}\Big)^{m-m_{q_i}} \\ &\quad - \frac{\binom{m-m_{q_i}}{2}}{T_{q_i-1}} \Big(\frac{W_{q_i-1}}{T_{q_i-1}}\Big)^{m_{q_i}} \Big(\frac{B_{q_i-1}}{T_{q_i-1}}\Big)^{m-m_{q_i}-1} + O\Big(\frac{1}{T_{q_i-1}^2}\Big)\Big) \\ &\quad \times \Big(1 + O\Big(\frac{1}{T_{q_i-1}}\Big)\Big) \end{aligned}$$

Recall that if $i \in \{1, \ldots, c_l\}$, $q_i > n + k$, then $\frac{W_{q_i}}{T_{q_i}} \overset{a.s}{=} x_1 + o\Big(\frac{\ln(n)}{\sqrt{n}}\Big)$.

$$\begin{aligned} w_{1,i} &\overset{a.s}{=} \binom{m}{m_{q_i}} \mathbb{E}\Big(\Gamma_{m_{q_i}}(q_i) |\xi_{q_j} = m_{q_j}, 1 \leqslant j \leqslant i-1\Big) \\ &\overset{a.s}{=} \binom{m}{m_{q_i}} x_1^{m_{q_i}} (1-x_1)^{(m-m_{q_i})} \Big(1 + o\Big(\frac{\ln(q_i - 1)}{\sqrt{q_i - 1}}\Big)\Big) \\ &\quad + O\Big(\frac{1}{q_i - 1}\Big)\Big)\Big(1 + O\Big(\frac{1}{q_i - 1}\Big)\Big) \\ &\overset{a.s}{=} \binom{m}{m_{q_i}} x_1^{m_{q_i}} (1-x_1)^{(m-m_{q_i})} \Big(1 + o\Big(\frac{\ln(q_i - 1)}{\sqrt{q_i - 1}}\Big)\Big) \\ &\overset{a.s}{=} \binom{m}{m_{q_i}} x_1^{m_{q_i}} (1-x_1)^{(m-m_{q_i})} \Big(1 + o\Big(\frac{\ln(n)}{\sqrt{n}}\Big)\Big). \end{aligned}$$

With a similar computation, we obtain $w_{h,i} = \binom{m}{m_{q_i}} x_1^{m_{q_i}} (1-x_1)^{(m-m_{q_i})} \Big(1 + o\Big(\frac{\ln(n)}{\sqrt{n}}\Big)\Big)$. It yields that

$$\left|w_{h,i} - w_{1,i}\right| = o\left(\frac{\ln(n)}{\sqrt{n}}\right). \tag{22}$$

As a conclusion, the inequality (20) becomes

$$\mathbb{C}ov(U, V) = o\left(\frac{\ln(n)}{\sqrt{n}}\right).$$

□

Theorem 4 *Let W_n be the number of white balls in the urn after n draws. Then W_n satisfies*

$$\frac{W_n - E(W_n)}{\sqrt{n}} \xrightarrow{\mathcal{D}} \mathcal{N}\left(0, \frac{a^2 m x_1 (1 - x_1)^2}{1 + x_1}\right). \tag{23}$$

Proof Let $\tilde{\xi}_i = a(\xi_i - \mathbb{E}(\xi_i))$ and $S_n = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \tilde{\xi}_i$. Using the weak dependence between the random variables $(X_i)_{1\leqslant i\leqslant n}$, we apply the Bernstein Method to prove the central limit theorem. This method consists in subdividing the interval $[0, n]$ into big and small intervals in a way that the considered sum in the big intervals, can be described as a sum of independent random variables and the same sum, in the small intervals, does not contributes to the asymptotic behavior. We can refer the reader to Lin and Lu (1996) for more details about this method. Consider the sequences p_n and q_n such that

$$p_n = [n^{\frac{3}{4}} \ln^2(n)] \quad \text{and} \quad q_n = \left[\frac{n^{\frac{1}{4}}}{\ln(n)}\right].$$

Let $k_n = [\frac{n}{p_n + q_n}]$ and

$$X_n = \frac{1}{\sqrt{n}} \sum_{i=1}^{k_n} \tilde{X}_i, \quad \text{and} \quad U_n = \frac{1}{\sqrt{n}} \sum_{i=1}^{k_n} \tilde{U}_i,$$

with

$$\tilde{X}_j = \sum_{i\in B_j} \tilde{\xi}_i, \quad \text{and} \quad \tilde{U}_j = \sum_{i\in B'_j} \tilde{\xi}_i,$$

where $B_j =](p_n + q_n)(j - 1), (p_n + q_n)(j - 1) + p_n[\cap \mathbb{N}$ is a subset of p_n successive integers from $\{1, \dots, n\}$ such that for $l \neq l'$, the distance between B_ℓ and $B_{\ell'}$ is at least q_n and B'_j is the block between B_j and B_{j+1}. Let U_{k_n} be the last sum of $\tilde{\xi}_i$ between the end of B_{k_n} and n. Let $(X_i^*)_{i\geqslant 0}$ be a sequence of independent random variables and independent of the sequence $(\tilde{X}_i)_{i\geqslant 0}$. Let

$$Y_n = \frac{1}{\sqrt{n}} \sum_{i=1}^{k_n} X_i^*,$$

such that for all $i \geqslant 1$, $X_i^* = \tilde{X}_i$ in distribution. Let Y be a random variable and $\Psi_Y(.)$ be the characteristic function of Y. Let N denote the centered normal random variable with variance $\sigma^2 = \frac{a^2 m x_1 (1-x_1)^2}{1+x_1}$. We have $S_n - N = (S_n - X_n) + (X_n - N)$. At first we have by Tchebychev inequality:

$$\forall \varepsilon > 0, \quad \mathbb{P}(|S_n - X_n| > \varepsilon) \leqslant \frac{\mathbb{V}ar(S_n - X_n)}{\varepsilon^2}. \tag{24}$$

Note that $\mathbb{E}\Big(|S_n - X_n|^2\Big) = \mathbb{E}(|U_n|^2)$, where

$$\mathbb{E}(U_n^2) = \mathbb{V}ar\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{k_n}\tilde{U}_i\right) = \frac{1}{n}\sum_{i=1}^{k_n}\mathbb{V}ar\left(\tilde{U}_i\right) + \frac{1}{n}\sum_{1\leqslant i<j\leqslant k_n}\mathbb{C}ov\left(\tilde{U}_i, \tilde{U}_j\right).$$

In view of Lemma 3 we have the following bounds:

$$\frac{1}{n}\sum_{j=1}^{k_n+1}\mathbb{V}ar(\tilde{U}_j) = \frac{1}{n}\sum_{j=1}^{k_n+1}\mathbb{V}ar\left(\sum_{j\in B_i'}\tilde{\xi}_j\right) = O\left(\frac{q_n^2 k_n}{n}\right) = O\left(\frac{1}{\ln(n)^4 n^{\frac{1}{4}}}\right). \tag{25}$$

Similarly, we have:

$$\begin{aligned}|\mathbb{C}ov(\tilde{U}_i, \tilde{U}_j)| &= |\sum_{k\in B_i'}\sum_{l\in B_j'}\mathbb{C}ov(\tilde{\xi}_i, \tilde{\xi}_j)| \\ &\leqslant q_n^2 m^2 \sup_{|i-j|>p_n,\alpha,\beta}|\mathbb{C}ov(1_{\{\xi_i=\alpha\}}, 1_{\{\xi_j=\beta\}})| \leqslant m^2 q_n^2 \alpha(p_n).\end{aligned}$$

Then,

$$\frac{1}{n}\sum_{1\leqslant i<j\leqslant k(n)}\mathbb{C}ov(\tilde{U}_i, \tilde{U}_j) = o\left(\frac{n^{-\frac{7}{8}}}{ln(n)^2}\right),$$

this proves that $S_n - X_n \xrightarrow{\mathbb{P}} 0$ when n tends to infinity. On the other hand, $\forall t \in]-1, 1[$ we have

$$|\Psi_{X_n}(t) - \Psi_N(t)| \leqslant |\Psi_{X_n} - \Psi_{Y_n}| + |\Psi_{Y_n} - \Psi_N(t)|. \tag{26}$$

For this, we use the following lemma:

Lemma 4 *We have $X_n - Y_n$ converges to 0 in distribution.*

Proof

$$\left|\mathbb{E}\left(e^{itX_n}\right) - \prod_{j=1}^{k_n}\mathbb{E}\left(e^{it\frac{\tilde{X}_j}{\sqrt{n}}}\right)\right| \leqslant \sum_{h=1}^{k_n}\left|\mathbb{E}\left(\prod_{j=h+1}^{k_n}e^{it\frac{\tilde{X}_j}{\sqrt{n}}}\right)\prod_{j=1}^{h}\mathbb{E}\left(e^{it\frac{\tilde{X}_j}{\sqrt{n}}}\right)\right.$$

$$
\begin{aligned}
&-\mathbb{E}\left(\prod_{j=h}^{k_n} e^{it\frac{\tilde{X}_j}{\sqrt{n}}}\right)\prod_{j=1}^{h-1}\mathbb{E}\left(e^{it\frac{\tilde{X}_j}{\sqrt{n}}}\right)\Bigg| \\
\leqslant &\sum_{h=1}^{k_n}\Bigg|\mathbb{E}\left(\prod_{j=h}^{k_n} e^{it\frac{\tilde{X}_j}{\sqrt{n}}}\right) \\
&-\mathbb{E}\left(\prod_{j=h+1}^{k_n} e^{it\frac{\tilde{X}_j}{\sqrt{n}}}\right)\mathbb{E}\left(e^{it\frac{\tilde{X}_h}{\sqrt{n}}}\right)\Bigg|
\end{aligned}
$$

We conclude the proof by using this lemma:

Lemma 5 (Lin and Lu 1996) *Let* $A \in \mathcal{G}_1^k$ *and* $B \in \mathcal{G}_{n+k}^{\infty}$ *such that* $|A| \leqslant C_1$ *and* $|B| \leqslant C_2$ *we have*

$$|\mathbb{C}ov(A, B)| \leqslant 4C_1C_2\alpha(n). \tag{27}$$

As a conclusion,

$$|\mathbb{E}\left(e^{itX_n}\right) - \prod_{j=1}^{k_n}\mathbb{E}\left(e^{it\frac{\tilde{X}_j}{\sqrt{n}}}\right)| \leqslant 4k_n\alpha(q_n) = o\left(\frac{n^{-\frac{1}{8}}}{\ln(n)^2}\right).$$

Finally, look at the limit of $\frac{1}{n}\sum_{i=1}^{k_n}\mathbb{E}\left(\tilde{X}_i^2\right)$.

$$
\begin{aligned}
\frac{1}{n}\sum_{i=1}^{k_n}\mathbb{E}\left(\tilde{X}_i^2\right) &= \frac{1}{n}\sum_{i=1}^{k_n}\mathbb{V}ar\left(\tilde{X}_i\right) \\
&= \frac{1}{n}\sum_{i=1}^{k_n}\mathbb{V}ar\left(\sum_{k\in B_i}\tilde{\xi}_k\right) \\
&= \frac{k_n p_n}{n}\frac{1}{k_n}\sum_{i=1}^{k_n}\frac{1}{p_n}\mathbb{V}ar\left(\sum_{k\in B_i}\tilde{\xi}_k\right).
\end{aligned}
$$

Recall that $\frac{1}{n}\mathbb{V}ar(\sum_{i=1}^{n}\tilde{\xi}_i)$ converges to $\frac{a^2mx_1(1-x_1)^2}{1+x_1}$. By Cesáro lemma we have $\frac{1}{k_n}\sum_{i=1}^{k_n}\frac{1}{p_n}\mathbb{V}ar(\sum_{k\in B_i}\tilde{\xi}_k)$ converges to the same limit. Indeed we have $\frac{k_n p_n}{n} = \left[\frac{p_n}{p_n+q_n}\right] = \left[\frac{1}{1+\frac{q_n}{p_n}}\right]$ which converges to 1. As a conclusion $\frac{1}{n}\sum_{i=1}^{k_n}\mathbb{E}\left(\tilde{X}_i^2\right) \xrightarrow{a.s} \frac{a^2mx_1(1-x_1)^2}{1+x_1}$ □

3 Self-reinforcement model

In this section, we will deal with unbalanced urn model submitted to the self-reinforcement rule. The dynamics of replacement are defined as follows:

$$W_{n+1} = W_n + a\xi_{n+1} \quad \text{and} \quad T_{n+1} = T_n + bm + (a-b)\xi_{n+1}. \tag{28}$$

3.1 Strong law of large number

In order to obtain the almost sure convergence of the proportion of white balls in the urn, we will relay as well on the stochastic approximation algorithm as in the previous section.

Lemma 6 *Let Z_n be the proportion of white balls after n draws in the urn submitted to the self-reinforcement rule. Then Z_n satisfies the following stochastic algorithm:*

$$Z_{n+1} - Z_n = \frac{1}{T_n} g(Z_n) + \frac{1}{T_n}\Big(\Delta M_{n+1}\Big), \tag{29}$$

where

$$g(x) = m(a-b)x(1-x) \quad \textit{and} \quad \Delta M_{n+1} = Y_{n+1} - \mathbb{E}(Y_{n+1}),$$

with $\quad Y_{n+1} = a\xi_{n+1} - Z_n(bm + (a-b)\xi_{n+1})$.

Proposition 4 *The proportion of white balls in the urn submitted to the self-reinforcement rule converges almost surely to* 0 *or* 1 *whenever $a < b$ or $a > b$.*

Proof The proportion of white balls in the urn satisfies Eq. (4), we will apply Theorem 1 with $\gamma_n = \frac{1}{T_n}$, $f \equiv g$ and $U_n = \Delta M_n$. Since Z_n is bounded by 1 we have:

$$\begin{aligned}|g(Z_n)| \leqslant m|a-b|, \quad |\Delta M_{n+1}| &= |(a-(a-b)Z_n)(\xi_{n+1} - mZ_n)| \\ &\leqslant 2m(a + |a-b|),\end{aligned}$$

and

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

As a result Z_n converges almost surely to the stable zeros of the function g which is 0 if $a < b$ and 1 if $a > b$. □

Throughout the rest of this paper, we will consider the case when $a < b$. The other case can be obtained analogously.

Proposition 5 *The total number of balls in the urn after n draws satisfies*

$$T_n = bmn + o(n^{\frac{a}{2b}+\frac{1}{2}+\delta}) \quad \textit{almost surely}. \tag{30}$$

where δ *is arbitrarily* $\in]0, 1 - \frac{a}{b}[$.

Proof Let $\varepsilon > 0$ and $k_n = [c^n]$ where $c > 1$. By Chebychev's inequality we have

$$\sum_{n\geqslant 0} \mathbb{P}\Big(|W_{k_n} - \mathbb{E}(W_{k_n})| > \varepsilon k_n^{\frac{a}{2b}+\frac{1}{2}+\delta}\Big) \leqslant \frac{1}{\varepsilon^2}\sum_{n\geqslant 0} \frac{\mathbb{V}ar(W_{k_n})}{k_n^{1+\frac{a}{b}+2\delta}} = \sum_{n\geqslant 0} o\left(\frac{1}{k_n^{\delta}}\right) < \infty.$$

By Borel-Cantelli lemma, when n tends to infinity, we have $\frac{W_{k_n}-\mathbb{E}(W_{k_n})}{k_n^{\frac{a}{2b}+\frac{1}{2}+\delta}} \xrightarrow{a.s} 0$. Let $k_n \leqslant k < k_{n+1}$, we have

$$\begin{aligned}
\frac{W_k - \mathbb{E}(W_k)}{k^{\frac{a}{2b}+\frac{1}{2}+\delta}} &\geqslant \frac{W_{k_n} - \mathbb{E}\left(W_{k_n}\right)}{k^{\frac{a}{2b}+\frac{1}{2}+\delta}} + \frac{\mathbb{E}\left(W_{k_n}\right) - \mathbb{E}\left(W_{k_{n+1}}\right)}{k^{\frac{a}{2b}+\frac{1}{2}+\delta}}\\
&\geqslant -\frac{|W_{k_n} - \mathbb{E}\left(W_{k_n}\right)|}{k_n^{\frac{a}{2b}+\frac{1}{2}+\delta}} - \frac{k_{n+1}^{\frac{a}{2b}+\frac{1}{2}+\delta}}{k_n^{\frac{a}{2b}+\frac{1}{2}+\delta}} \frac{\mathbb{E}\left(W_{k_{n+1}}\right) - \mathbb{E}\left(W_{k_n}\right)}{k_{n+1}^{\frac{a}{2b}+\frac{1}{2}+\delta}}\\
&\geqslant -\varepsilon - c^{\frac{a}{b}+1+2\delta} \frac{C_1\left(k_{n+1}^{\frac{a}{b}+o(1)} - k_n^{\frac{a}{b}+o(1)}\right)}{k_{n+1}^{\frac{a}{2b}+\frac{1}{2}+\delta}}\\
&\geqslant -\varepsilon - \frac{C_1 c^{\frac{a}{b}+1+2\delta}}{c^{n\delta+o(n)}}\left(1 - c^{-\frac{2a}{b}+o(1)}\right)\\
&\geqslant -\varepsilon - \left(1 - c^{-\frac{2a}{b}+o(1)}\right)\varepsilon \text{ for } n \text{ too large.}
\end{aligned}$$

Making $c \searrow 1$ we obtain $\liminf \frac{W_k-\mathbb{E}(W_k)}{k^{\frac{a}{2b}+\frac{1}{2}+\delta}} \geqslant 0$. Similarly, for the lim sup we have

$$\begin{aligned}
\frac{W_k - \mathbb{E}(W_k)}{k^{\frac{a}{2b}+\frac{1}{2}+\delta}} &\leqslant \frac{k_{n+1}^{\frac{a}{2b}+\frac{1}{2}+\delta}}{k_n^{\frac{a}{2b}+\frac{1}{2}+\delta}} \frac{|W_{k_{n+1}} - \mathbb{E}\left(W_{k_{n+1}}\right)|}{k_{n+1}^{\frac{a}{2b}+\frac{1}{2}+\delta}} + \frac{\mathbb{E}\left(W_{k_{n+1}}\right) - \mathbb{E}\left(W_{k_n}\right)}{k_n^{\frac{a}{2b}+\frac{1}{2}+\delta}}\\
&\leqslant c^{\frac{a}{b}+1+2\delta}\varepsilon + C_1\left(c^{2\frac{a}{b}+o(1)} - 1\right)\frac{1}{c^{(n-1)\delta+o(n)}}
\end{aligned}$$

Making $c \searrow 1$ we have the upper bound $\limsup \frac{W_k-\mathbb{E}(W_k)}{k^{\frac{a}{2b}+\frac{1}{2}+\delta}} \leqslant \varepsilon$. Recall that we have the relation $aB_n + bW_n = abmn + aB_0 + bW_0$. Thus, $T_n = bmn - \frac{b}{b-a}W_n + \frac{aB_0+bW_0}{b}$ which concludes the proof. □

Proposition 6 *The mean and the variance of the white balls in the urn after n draws satisfies*

$$\mathbb{E}(W_n) = m\frac{W_0}{T_0}\exp\left(\frac{a}{b}\gamma + \sum_{j=2}^{\infty}\frac{(-1)^{j-1}}{j}\left(\frac{a}{b}\right)^j \zeta(j)\right) n^{\frac{a}{b}}\Big(1 + o\Big(n^{\frac{a}{2b}-\frac{1}{2}+\delta}\Big)\Big), \tag{31}$$

$$\text{and} \tag{32}$$

$$\mathbb{V}ar(W_n) = \left(C_2 V_1 + \frac{a^2 C_1}{b}\zeta\Big(1+\frac{a}{b}\Big) - \frac{a^2\pi^2 C_1^2}{6b^2 m}\right) n^{\frac{2a}{b}}\Big(1 + o(n^{\frac{a}{2b}-\frac{1}{2}+\delta})\Big), \tag{33}$$

where $C_1 = m\frac{W_0}{T_0}\exp\left(\frac{a}{b}\gamma + \sum_{j=2}^{\infty}\frac{(-1)^{j-1}}{j}\left(\frac{a}{b}\right)^j\zeta(j)\right)$ *and* $C_2 = \exp\Big[\frac{2a}{b}\gamma - \frac{\pi^2 a^2}{6b^2 m} - \sum_{j=2}^{\infty}\frac{1}{j}\left(\frac{a^2}{b^2 m}\right)^j \sum_{h=0}^{j}\binom{j}{h}(-1)^h\Big(\left(\frac{a}{2bm}\right)^h\zeta(2j-h)\Big)\Big]$.

Proof

$$\begin{aligned}
\mathbb{E}(W_n) &= \mathbb{E}(W_1)\prod_{k=1}^{n-1}\Big(1+\frac{a}{bk}(1+o(k^{\frac{a}{2b}-\frac{1}{2}+\delta}))\Big)\\
&= \mathbb{E}(W_1)\exp\left(\sum_{k=1}^{n-1}\ln\Big(1+\frac{a}{bk}\Big(1+o\Big(k^{\frac{a}{2b}-\frac{1}{2}+\delta}\Big)\Big)\Big)\right)\\
&= \mathbb{E}(W_1)\exp\Big[\frac{a}{b}\Big(\ln(n)+\gamma+O\Big(\frac{1}{n}\Big)\Big) + \sum_{j=2}^{\infty}\frac{(-1)^{j-1}}{j}\Big(\frac{a}{b}\Big)^j\Big(\zeta(j)\\
&\quad +O\Big(\frac{1}{n}\Big)\Big)\Big(1+o(n^{\frac{a}{2b}-\frac{1}{2}+\delta})\Big)\Big]\\
&= m\frac{W_0}{T_0}\exp\left(\frac{a}{b}\gamma + \sum_{j=2}^{\infty}\frac{(-1)^{j-1}}{j}\Big(\frac{a}{b}\Big)^j\zeta(j)\right) n^{\frac{a}{b}}\Big(1+o\Big(n^{\frac{a}{2b}-\frac{1}{2}+\delta}\Big)\Big).
\end{aligned}$$

Let $C_1 = m\frac{W_0}{T_0}\exp\left(\frac{a}{b}\gamma + \sum_{j=2}^{\infty}\frac{(-1)^{j-1}}{j}\left(\frac{a}{b}\right)^j\zeta(j)\right)$.

$$\begin{aligned}
\mathbb{V}ar(W_{n+1}) &= \mathbb{V}ar(W_n) + a^2 m E\left(\frac{W_n}{T_n} - \frac{W_n^2}{T_n^2}\right)\Big(1+O\Big(\frac{1}{n}\Big)\Big) + 2a\mathbb{C}ov(W_n,\xi_{n+1})\\
&= \mathbb{V}ar(W_n) + \frac{a^2}{bn}E(W_n)\Big(1+o\Big(n^{\frac{a}{2b}-\frac{1}{2}+\delta}\Big)\Big)\\
&\quad -\frac{a^2}{b^2 m n^2}E(W_n^2)\Big(1+o\Big(n^{\frac{a}{2b}-\frac{1}{2}+\delta}\Big)\Big)\\
&\quad +\frac{2a}{bn}\mathbb{V}ar(W_n)\Big(1+o(n^{\frac{a}{2b}-\frac{1}{2}+\delta})\Big)\\
&= \Big(1+\Big(\frac{2a}{bn}-\frac{a^2}{b^2 m n^2}\Big)\Big(1+o\Big(n^{\frac{a}{2b}-\frac{1}{2}+\delta}\Big)\Big)\Big)\mathbb{V}ar(W_n)\\
&\quad +\Big(\frac{a^2 C_1}{bn}n^{\frac{a}{b}} - \frac{a^2 C_1^2}{b^2 m n^2}n^{\frac{2a}{b}}\Big)(1+o(n^{\frac{a}{2b}-\frac{1}{2}+\delta}))\\
&= \left(\prod_{k=1}^{n-1} w_k\right)\Big(\mathbb{V}ar(W_1) + \sum_{k=1}^{n-1}\frac{\theta_k}{\prod_{j=1}^{k-1} w_j}\Big),
\end{aligned}$$

where

$$w_n = 1+\Big(\frac{2a}{bn}-\frac{a^2}{b^2mn^2}\Big)\Big(1+o\Big(n^{\frac{a}{2b}-\frac{1}{2}+\delta}\Big)\Big) \quad \text{and} \quad \theta_n = \Big(\frac{a^2C_1}{bn}n^{\frac{a}{b}}-\frac{a^2C_1^2}{b^2mn^2}n^{\frac{2a}{b}}\Big)(1+o(n^{\frac{a}{2b}-\frac{1}{2}+\delta})).$$

First consider $Q_n := \ln\Big(\prod_{k=1}^{n-1} w_k\Big)$:

$$\begin{aligned}
Q_n &= \sum_{k=1}^{n-1}\ln\Big(1+\Big(\frac{2a}{bn}-\frac{a^2}{b^2mn^2}\Big)\Big(1+o\Big(n^{\frac{a}{2b}-\frac{1}{2}+\delta}\Big)\Big)\Big)\\
&= \sum_{k=1}^{n-1}\sum_{j=1}^{\infty}\frac{(-1)^{j-1}}{j}\Big(\frac{2a}{bn}-\frac{a^2}{b^2mn^2}\Big)^j\Big(1+o\Big(n^{\frac{a}{2b}-\frac{1}{2}+\delta}\Big)\Big)\\
&= \sum_{k=1}^{n-1}\Big(\frac{2a}{bn}-\frac{a^2}{b^2mn^2}\Big)\Big(1+o\Big(n^{\frac{a}{2b}-\frac{1}{2}+\delta}\Big)\Big)\\
&\quad -\sum_{k=1}^{n-1}\sum_{j=2}^{\infty}\frac{1}{j}\sum_{h=0}^{j}\binom{j}{h}(-1)^h\Big(\frac{2a}{bk}\Big)^h(\frac{a^2}{b^2mk^2})^{j-h}\Big(1+o\Big(n^{\frac{a}{2b}-\frac{1}{2}+\delta}\Big)\Big)\\
&= \Big(\frac{2a}{b}(\ln(n)+\gamma)-\frac{\pi^2a^2}{6b^2m}+O\Big(\frac{1}{n}\Big)\Big)\Big(1+o\Big(n^{\frac{a}{2b}-\frac{1}{2}+\delta}\Big)\Big)\\
&\quad -\sum_{j=2}^{\infty}\frac{1}{j}\Big(\frac{a^2}{b^2m}\Big)^j\sum_{h=0}^{j}\binom{j}{h}(-1)^h\Big(\frac{a}{2bm}\Big)^h\zeta(2j-h)\Big(1\\
&\quad +O\Big(\frac{1}{n^{2j-h}}\Big)\Big)\Big(1+o(n^{\frac{a}{2b}-\frac{1}{2}+\delta})\Big).
\end{aligned}$$

We obtain, for n too large $\Big(\prod_{k=1}^{n-1} w_k\Big) = C_2 n^{\frac{a}{2b}}(1+o(n^{\frac{a}{2b}-\frac{1}{2}+\delta}))$ where C_2 is given by

$$C_2 = \exp\Big(\frac{2a}{b}\gamma-\frac{\pi^2a^2}{6b^2m}-\sum_{j=2}^{\infty}\frac{1}{j}\Big(\frac{a^2}{b^2m}\Big)^j\sum_{h=0}^{j}\binom{j}{h}(-1)^h\Big(\Big(\frac{a}{2bm}\Big)^h\zeta(2j-h)\Big)\Big). \tag{34}$$

This yields the results. For S_n we have $S_n = \sum_{k=1}^{n-1}\frac{\theta_k}{\prod_{j=1}^{k-1}w_j}$:

$$\begin{aligned}
S_n &= \Big(\sum_{k=1}^{n-1}\frac{a^2C_1}{bC_2k^{\frac{a}{b}+1}}-\sum_{k=1}^{n-1}\frac{a^2C_1^2}{b^2mC_2k^2}\Big)\Big(1+o(n^{\frac{a}{2b}-\frac{1}{2}+\delta})\Big)\\
&= \Big(\frac{a^2C_1}{bC_2}\zeta(1+\frac{a}{b})\Big)\Big(1+O\Big(\frac{1}{n^{\frac{a}{b}}}\Big)\Big)-\frac{\pi^2a^2C_1^2}{6bC_2}\Big(1+O\Big(\frac{1}{n}\Big)\Big)(1+o(n^{\frac{a}{2b}-\frac{1}{2}+\delta})).
\end{aligned}$$

□

Remark 3 In the case where $a = b$, the total number of balls in the urn is deterministic, Chen and Kuba (2013) obtained the moments of high orders in a recurrence form.

3.2 Limit in distribution

Theorem 5 *There exists a random variable W_∞ such that the number of white balls in the generalized urn submitted to the self-reinforcement rule satisfies, as n tends to infinity*

$$\frac{W_n}{n^{\frac{a}{b}}} \longrightarrow W_\infty \text{ almost surely.} \tag{35}$$

The mean and the variance of W_∞ are given by

$$E(W_\infty) = m\frac{W_0}{T_0}\exp\left(\frac{a}{b}\gamma + \sum_{j=2}^{\infty}\frac{(-1)^{j-1}}{j}\left(\frac{2a}{b}\right)^j\zeta(j)\right) \tag{36}$$

and

$$\mathbb{V}ar(W_\infty) = C_2V_1 + \frac{a^2C_1}{b}\zeta\left(1+\frac{a}{b}\right) - \frac{a^2\pi^2C_1^2}{6b^2m}, \tag{37}$$

where $V_1 = m\frac{W_0}{T_0}\left(1-\frac{W_0}{T_0}\right)\frac{T_0-m}{T_0-1}$, $C_1 = E(W_\infty)$, *and* $C_2 = \exp\left(\frac{2a}{b}\gamma - \frac{\pi^2a^2}{6b^2m} - \sum_{j=2}^{\infty}\frac{1}{j}\left(\frac{a^2}{b^2m}\right)^j\sum_{h=0}^{j}\binom{j}{h}(-1)^h\left(\frac{a}{2bm}\right)^h\zeta(2j-h)\right)$. *Moreover, the distribution of W_∞ is absolutely continuous.*

Proof The sequence $M_n = \prod_{k=0}^{n-1}\left(1+\frac{am}{T_k}\right)^{-1}W_n$ is a martingale adapted to the filtration $(\mathcal{F}_n)_{n\geqslant 0}$. In view of Eq. (30), there exists a positive constant α such that, for some $\delta\in]0, 1-\frac{a}{b}[$, we have

$$\prod_{k=0}^{n-1}\left(1+\frac{am}{T_k}\right)^{-1} = \frac{e^\alpha}{n^{\frac{a}{b}}}(1+O(n^{\frac{a}{b}+\delta-1}\vee n^{\delta-\frac{1}{2}})), \quad \text{almost surely.}$$

To prove that W_∞ has an absolutely continuous distribution we will follow the proof given by Chen and Wei (2005). Define the increasing events Ω_ℓ by

$$\Omega_\ell = \{W_\ell \geqslant am, \text{ and } B_\ell \geqslant bm\}.$$

Since $P(\cup_{\ell\geq 1}\Omega_\ell) = 1$, it is sufficient to prove that W_∞ has a density on $\Omega_{\ell,j} = \Omega_\ell \cap \{W_\ell = j\}$.

Lemma 7 (Chen and Wei 2005) *There exists a positive constant c such that for every $n \geqslant \ell+1$ and $0 \leq k \leqslant m(n+1)$ and $am \leqslant j \leqslant amn$*

$$\sum_{i=0}^{m} P(W_{n+1} = j+ak|W_n = j+a(K-i)) \leqslant 1-\frac{1}{n}+\frac{c}{n^2} \tag{38}$$

Lemma 8 (Chen and Wei 2005) *There exists a positive constant C depending only on* ℓ *such that for every* $\ell \geqslant 1$ *and* $n \geqslant \ell + 1$ *and* $am \leqslant j \leqslant amn$: *we have*

$$\max_{0 \leqslant k \leqslant m(n+1)} P(W_{n+1} = j + ak | W_\ell = j) \leqslant \frac{C(\ell)}{n}. \tag{39}$$

Let $\varepsilon > 0$ and $\delta = \frac{\varepsilon}{C(\ell)}$, and set $x_1 < x_1' \leqslant x_2 < x_2' \leqslant \cdots \leqslant x_r < x_r'$ such that $\sum_{i=1}^r |x_i' - x_i| \leqslant \delta$. Using Fatou's lemma we have

$$\begin{aligned}
\sum_{i=1}^r P(\{x_i \leqslant W_\infty \leqslant x'\} \cap \Omega_{\ell,j}) &\leqslant \sum_{i=1}^r \liminf P(x_i \leqslant \frac{W_n}{n} \leqslant x_i' | W_\ell = j) P(\Omega_{\ell,j}) \\
&\leqslant \sum_{i=1}^r \liminf \left(((x_i' - x_i)n + 1) \frac{C(\ell)}{n} \right) \\
&\leqslant \sum_{i=1}^r (x_i' - x_i) C(\ell) = \varepsilon.
\end{aligned}$$

□

Example 1 The case where $m = 1$ corresponds to the model studied by Janson (2006) where he proved that when $a < b$ the number of white balls satisfies the following convergence:

$$\frac{bn - B_n}{n^{\frac{a}{b}}} \xrightarrow{\mathcal{L}} bUV^{-\frac{a}{b}},$$

where $U \stackrel{D}{=} \Gamma\left(\frac{W_0}{a}, 1\right)$ and $V \stackrel{D}{=} \Gamma\left(\frac{B_0}{b}, 1\right)$.

Acknowledgements The first author is grateful to the King Saud University, Deanship of Scientific Research, College of Science Research Center. The authors are grateful to anonymous referees for their valuable comments and corrections that improved the presentation of this paper.

References

Athreya, K. B., Ney, P. E. (1972). *Branching Processes*. Berlin: Springer.

Baggchi, A., Paul, A. K. (1985). Asymptotic normality in the generalized Pólya-Eggenberger urn model with application to computer data structure. *SIAM Journal on Algebraic and Discrete Methods*, *6*(3), 394–405.

Benveniste, A., Métivier, M., Priouret, P. (1990). *Adaptive algorithms and stochastic approximations*. Berlin: Springer.

Chauvin, B., Pouyanne, N., Sahnoun, R. (2011). Limit distributions for large Pólya urns. *Annals of Applied Probability*, *21*(1), 1–32.

Chen, M. R., Kuba, M. (2013). On generalized Polya urn models. *Journal of Applied Probability Theory*, *50*(4), 1169–1186.

Chen, M. R., Wei, C. Z. (2005). A new urn model. *Journal of Applied Probability*, *42*(4), 964–976.

Duflo, M. (1997). *Random iterative models*. Berlin: Springer.

Eggenberger, F., & Polya, G. (1923). Über die Statistik verketteter Vorgänge. *ZAMM - Journal of Applied Mathematics and Mechanics*, *3*(4), 279–289.

Fazekas, I., Klesov, O. (2001). On a General Approach to the Strong Laws of Large Numbers. *Theory of Probability and its Application*, *45*(3), 436–449.

Flajolet, P., Gabarro, J., Pekari, H. (2005). Analytic urns. *The Annal of Probability*, *33*(3), 1200–1233.

Janson, S. (2006). Limit theorems for triangular urn schemes. *Probability Theory and Related Fields*, *134*(3), 417–452.

Konzem, S., Mahmoud, H. (2016). Characterization and enumeration of certain classes of Tenable Polya urn schemes growing under drawing multisets of balls. *Methodology and Computing in Applied Probability*, *18*(2), 359–375.

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

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

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

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

Longnecker, M., Serfling, R. J. (1977). General moment and probability inequalities for the maximum partial sum. *Acta Mathematica*, *30*, 129–133.

Pagés, G., Laruelle, S. (2015). Randomized urns models revisited using stochastic approximation. *Annals of Applied Probability*, *23*(4), 1409–1436.

Renlund, H. (2010). Generalized Polya urns via stochastic approximation. arXiv:1002.3716v1.

Renlund, H. (2011). Limit theorem for stochastic approxiamtion algorithm. arXiv:1102.4741v1.

Tsukiji, T., Mahmoud, H. (2001). A limit law for outputs in random circuits. *Algorithmica*, *31*(3), 403–412.