Almost Sure Convergence of Randomized Urn Models with Finite Mean

By Ujan Gangopadhyay and Krishanu Maulik

University of Southern California and Indian Statistical Institute

We consider a randomized urn model containing objects of finitely many colors in this article. The replacement matrices are allowed to be random, subject to the minimum conditions that the mechanism to choose color and the replacement matrix at each step are conditionally independent given the past, as well as, the conditional expectations of the replacement matrices are close to a (possibly random) irreducible (and hence positive recurrent) matrix. We obtain almost sure convergence of the configuration vector, the proportion vector and the count vector under finite first moment condition alone. The convergence is shown to be $L^1$ as well. We show that first moment assumption is sufficient when the replacement matrix sequence is i.i.d. and independent of the past choices of the color. This significantly improves the similar results for urn models obtained by Athreya and Ney (1972), by weakening the moment assumptions on replacement matrices from $L \log L$ to $L^1$. For more general adaptive sequence of replacement matrix, a little more than $L \log L$ condition is required.

Dedicated to Krishna B. Athreya on the occasion of his 80th birth year.

1. Introduction. Urn models have provided a simple and useful model for studying different phenomena since the early works of Pólya and Eggenberger (1923) and Pólya (1930). The model starts with certain quantities of objects of different colors - we shall consider number of colors to be finite. A color is chosen at random, and further objects of each color is added to the urn, where the additional amount of each color added will depend on the color chosen. The additional amount of each color can be summarized in a square matrix, called replacement matrix, indexed by the colors, where the row corresponding to the color chosen will provide the amount of additional objects to be added when that color is chosen. We shall allow the amount added to be nonnegative real valued.

MSC 2010 subject classifications: Primary 62L20; secondary 60F15, 60G42

Keywords and phrases: urn model, random replacement matrix, balanced replacement matrix, irreducibility, stochastic approximation, random step size, random drift, uniform integrability, Lotka-Volterra differential equation
Classically, the objects are taken as balls and thus amount considered become counts and are nonnegative integer valued. Originally, Pólya considered addition of balls only of the color drawn and number of balls added was same for all colors drawn, leading to a replacement matrix which is a multiple of identity matrix. In this case, the vector of proportion of each color converges almost surely to a Dirichlet random vector, whose parameters will depend on the initial count of each color. However, a very different kind of replacement matrix, namely irreducible one, is also considered in the literature. A square matrix $A$ with nonnegative entries is called irreducible if, for any $i, j$ there exists a positive integer $N \equiv N_{ij}$ such that $A^{(N)}_{ij} > 0$, where $A^{(N)}_{ij}$ denotes the $(i, j)$-th entry of $A^N$. When $A$ has all row sums as 1, this notion of irreducibility coincides with the usual one considered in the theory of Markov chains. By Perron Frobenius theory, the eigenvalue of an irreducible matrix having largest real part is simple and positive. This eigenvalue is called the dominant eigenvalue. All other eigenvalues will be equal to or smaller than the dominant eigenvalue in modulus. If the matrix is balanced, namely, all the row sums are equal, then the dominant eigenvalue is the common row sum. Otherwise, the dominant eigenvalue lies strictly between the largest and the smallest row sums. Also corresponding to the dominant eigenvalue, there exists left and right eigenvectors which have all coordinates positive. In this case, the vector of proportion of each color converges almost surely to a deterministic vector which is the left eigenvector, normalized to a probability vector, of the replacement matrix corresponding to the dominant eigenvalue. In Athreya and Ney (1972), the replacement matrices were allowed to be an i.i.d. sequence with irreducible mean matrix. Using branching process techniques, under additional $L \log + L$ moment conditions on the entries of the replacement matrices, Athreya and Ney (1972), cf. Chapter V.9.3, showed the almost sure convergence to the (normalized to probability) left eigenvector of the mean matrix corresponding to its dominant eigenvalue. It may be noted that we weaken the moment condition in the similar setup significantly and show that first moment alone suffices, cf. Corollary 3.6.

However, applications to clinical trials showed that the i.i.d. assumption on the replacement matrices can be too restrictive. The replacement matrices may be dependent on the entire past. Thus Bai and Hu (2005) and Laruelle and Pagès (2013) consider a setup where the replacement matrix at any step is only assumed to be independent of the drawing mechanism of that step given its entire history. While Bai and Hu (2005) used martingale techniques, Laruelle and Pagès (2013) used stochastic approximation. Both of them assumed the conditional mean of the replacement matrices...
to concentrate around an irreducible, balanced matrix and obtained almost sure convergence and central limit type behavior of the proportion vector under second or higher moment conditions. Zhang (2012), using martingale based techniques, removed the balanced assumption and relaxed the mode of convergence of the conditional mean replacement matrices. He also proved the almost sure convergence under $L(\log_+ L)^p$ condition, for some $p > 1$; and, under $L \log_+ L$ condition, but only for i.i.d. replacement matrices. Note that the latter result is essentially the result obtained by Athreya and Ney (1972). Further, convergence in probability has been obtained in Gangopadhyay and Maulik (2019) under $L^1$ or first moment condition alone.

In this article, we significantly relax the assumptions to obtain almost sure convergence results applying the method of stochastic approximation. While we obtain almost sure convergence under $L(\log_+ L)^p$ condition for some $p > 1$, as in Zhang (2012), we show the moment condition can be further relaxed; and only $L \log_+ L$ condition will suffice under appropriate majorization condition on the norms of the replacement matrices. The additional moments in Zhang (2012) or the majorization condition provide the required uniform integrability. As mentioned before, in case of i.i.d. replacement matrices, almost sure convergence was proved by Athreya and Ney (1972) and Zhang (2012) under $L \log_+ L$ condition. In this article, we establish almost sure convergence result for i.i.d. replacement matrices under first moment conditions alone, please see Remark 2.11 and Assumption 2.2 therein, as well as Corollary 3.6. As in Gangopadhyay and Maulik (2019); Renlund (2010, 2011), we use random step sizes instead of deterministic ones considered in Zhang (2016). This leads to a Lotka-Volterra type first order quadratic differential equation. It has been shown to have an explicit unique solution in Gangopadhyay and Maulik (2019).

In Section 2, we collect some notations and state the model with the assumptions. We discuss the relevance of and sufficient conditions for the assumptions. In this section, we also recall the stochastic approximation result and collect some other useful results. In particular, we study convergence of appropriate martingale sequences under bounds either given by stochastic dominance or given by functions belonging to a general class which includes $x(\log_+ x)^p$, $x \log_+ x(\log_+ \log_+ x)^p$ for $p > 1$. We also quote the necessary results on Stochastic Approximation and prove some sufficient conditions appropriate for our work. In Section 3, we check the relevant conditions for stochastic approximation and prove the main Theorem 3.5 and some corollaries. In particular, in Corollary 3.6, we extend the result in Athreya and Ney (1972) and Zhang (2012) and prove it under weaker $L^1$
assumptions rather than $L \log L$ conditions.

2. Urn Model and Stochastic Approximation. In this Section, we first collect some notations. In three subsequent subsections, we state the model, together with a discussion on the assumptions; gather some useful results; and provide a quick overview of stochastic approximation with the relevant result. In the following, all the vectors considered are row vectors, while for a row vector $x$, its transpose will be denoted by $x^T$. The vector with all components 1 will be denoted by $1$, while the dimension will be clear from the context. Also, for an event $A$, we shall denote its indicator function by $1_A$. For a matrix $A$, define the norm $\rho(A) := \max_i \sum_j |A_{ij}|$.

2.1. Urn Model. Consider an urn model with finitely many $K$ colors indexed by $\{1, 2, \ldots, K\}$. The composition vector after $n$-th trial will be denoted by $C_n = (C_n1, \ldots, C_nK)$. We shall denote the total content of the urn after $n$-th trial by $S_n = \sum_{i=1}^K C_{ni}$. The replacement matrix for $n$-th trial will be a possibly random, but non-negative $K \times K$ matrix $R_n$. For $n$-th trial, $\chi_n$ will be the $K$-dimensional indicator vector of the color chosen in that trial, which will take value $e_i$, the $i$-th coordinate vector in $\mathbb{R}^K$ if color $i$ is chosen in $n$-th trial. Therefore for $n \geq 1$, the urn composition evolves as:

$$C_n = C_{n-1} + \chi_n R_n.$$  

For $n \geq 1$, let $\mathcal{F}_n$ be the sigma-field containing the entire information till time $n$. In particular, $\mathcal{F}_n$ will be the sigma-field generated by $C_0$, $(R_m)_{m=1}^n$ and $(\chi_m)_{m=1}^n$. We also need to consider the conditional expectation of the replacement matrices, denoted as $H_{n-1} := \mathbb{E}(R_n |\mathcal{F}_{n-1})$ and its truncated version, denoted as $\tilde{H}_{n-1} := \mathbb{E}(R_n 1_{\{\rho(R_n) \leq n\}} |\mathcal{F}_{n-1})$, called the generating matrices and the truncated generating matrices respectively.

We make the following assumptions on choice of colors and reinforcement matrices.

Assumption 2.1. The adapted sequence $((\chi_n, R_n), \mathcal{F}_n)$ has distribution, which satisfies:

(i) The initial configuration $C_0$ is nonzero with nonnegative entries and finite mean.

(ii) For all $i$, $\mathbb{P}(\chi_n = e_i | C_0, (R_m)_{m=1}^{n-1}, (\chi_m)_{m=1}^{n-1}) = C_{n-1,i}/S_{n-1}$. 
(iii) Given the past information \( \mathcal{F}_{n-1}, \chi_n \) and \( R_n \) are conditionally independent.

(iv) The tail of the sequence \( \rho(R_n) \) satisfy one of the following uniform integrability like conditions:

(iv-a) The unconditional distributions of \( \rho(R_n) \) are majorized: there exists a number \( c \in (0, \infty) \) and a random variable \( \tilde{R} \) with finite expectation such that, for all \( x > 0, \)

\[
P(\rho(R_n) > x) \leq c P(\tilde{R} > x).
\]

(iv-b) For some nonnegative function \( \phi \) on \([0, \infty)\) which is eventually positive and nondecreasing, satisfies \( \sum 1/(n\phi(n)) < \infty \) and further satisfies \( x/\phi(x) \) to be either bounded or eventually monotone nondecreasing, we have

\[
\sup_n \mathbb{E}(\rho(R_n) \phi(\rho(R_n)) | \mathcal{F}_{n-1}) < \infty
\]

or,

\[
\sup_n \mathbb{E}(\rho(R_n) \phi(\rho(R_n))) < \infty.
\]

(v) There exists a (possibly random) matrix \( H \) with nonnegative entries, which is almost surely irreducible, such that the truncated generating matrices \( \tilde{H}_n \) converge to \( H \) in Cesaro sense in the operator norm almost surely, i.e.,

\[
\frac{1}{n} \sum_{k=0}^{n-1} \rho(\tilde{H}_k - H) \to 0 \text{ almost surely.}
\]

**Remark 2.1.** Observe that \( \mathbb{E}(\chi_n | \mathcal{F}_{n-1}) = C_{n-1}/S_{n-1} \).

**Remark 2.2.** Under Assumption 2.1(iv), it is easy to see that, for all \( n, i, j, R_{nij} \) have finite mean.

**Remark 2.3.** To construct the sequences of random variables \( \chi_n \) and \( R_n \) as conditionally independent given the past and with required conditional distribution for \( \chi_n \), we may consider an i.i.d. Uniform(0,1) sequence \( (U_n) \) independent of the sequence \( (R_n) \). Then defining

\[
\chi_n = \sum_{i=1}^{d} e_i 1_{[U_n \in (\sum_{j=1}^{i-1} C_{n-1,j}/S_{n-1}, \sum_{j=1}^{i} C_{n-1,j}/S_{n-1})]},
\]

will construct the required sequence \( \chi_n \).
Remark 2.4. Assumption 2.1(iv-a) or (2.3) of Assumption 2.1(iv-b) holds, for example, if \((\rho(R_n))\) have same marginal distribution.

Remark 2.5. One of the useful choices for \(\phi(x) = (\log_+ x)^p\) for some \(p > 1\), as has been considered by Zhang (2012). Other choices can include \(\phi(x) = \log_+ x(\log_+ \log_+ x)^p\) for some \(p > 1\).

Remark 2.6. Each of Assumption 2.1(iv-a) and condition (2.3) in Assumption 2.1(iv-b) implies uniform integrability of \((\rho(R_n))\). Then the sequence \(\{R_n 1_{[\rho(R_n) \leq n]}\}\) is also uniformly integrable. Hence under either of these conditions, by Lemma 3.2 of Gangopadhyay and Maulik (2019), the convergence in Assumption 2.1(v) is also in \(L^1\).

Remark 2.7. Besides assuming \(R_n\) to be conditionally independent from \(\chi_n\) given the past, we also need the conditional expectations of \(R_n\) to be close to a limiting matrix \(H\) at least in some weak sense. Such assumptions have been considered by Bai and Hu (2005); Gangopadhyay and Maulik (2019); Laruelle and Pagès (2013); Zhang (2012, 2016). See Gangopadhyay and Maulik (2019) for further discussion on such assumptions.

Remark 2.8. It follows from Perron-Frobenius theorem that the eigenvalue of \(H\) with largest real part, \(\lambda_H\), is simple, real and positive. All other eigenvalues are less than or equal to \(\lambda_H\) in modulus and have strictly smaller real part. Further, there exists unique left eigenvector \(\pi_H\) corresponding to \(\lambda_H\) such that \(\pi_H\) has all coordinates strictly positive and is normalized to be a probability vector. The eigenvalue \(\lambda_H\) further satisfies \(0 < \sigma(H) \leq \lambda_H \leq \rho(H) < \infty\). The inequalities are strict, if the matrix \(H\) is not balanced, that is \(\sigma(H) < \rho(H)\).

Remark 2.9. We provide a sufficient condition for Assumption 2.1(v) in terms of nontruncated generating matrices, which is more usual, and a tail condition. Assumption 2.1(v) holds if both the conditions below hold:

(a) There exists a (possibly random) matrix \(H\) with nonnegative entries, which is almost surely irreducible, such that the generating matrices \(H_n\) converge to \(H\) in Cesaro sense in the operator norm almost surely:

\[
\frac{1}{n} \sum_{k=0}^{n-1} \rho(H_k - H) \to 0 \text{ almost surely.}
\]
(b) Negligibility of the tail conditional expectation:

\[ \frac{1}{n} \sum_{k=1}^{n} \mathbb{E} \left( \rho(R_k) \mathbb{1}_{[\rho(R_k) > k]} | F_{k-1} \right) \rightarrow 0 \text{ almost surely.} \]

Next we consider some sufficient condition for (2.6) to hold.

**Proposition 2.1.** Along with Assumption 2.1(i)--(iii), each of the following conditions is sufficient for (2.6) to hold:

(i) Assumption 2.1(iv-a) and, for each \( n \), \( R_n \) is independent of \( F_{n-1} \).

(ii) Assumption 2.1(iv-a) and the majorizing random variable \( R \) therein satisfies \( \mathbb{E}(R \log R) < \infty \).

(iii) Assumption 2.1(iv-a) and the conditional distributions of \( \rho(R_n) \) given \( F_{n-1} \) are majorized, that is, there exists number \( c' \in (0, \infty) \) and a random variable \( X \) with finite mean such that, for all large \( n \) and for all \( x > 0 \), we have

\[ \mathbb{P}( \rho(R_n) > x | F_{n-1} ) \leq c' \mathbb{P}( X > x ). \]

(iv) Assumption 2.1(iv-b).

**Proof.** We check each of the conditions one at a time.

(i) By independence, the left side of (2.6) becomes

\[ \frac{1}{n} \sum_{k=1}^{n} \mathbb{E} \left( \rho(R_k) \mathbb{1}_{[\rho(R_k) > k]} \right) \leq c \frac{1}{n} \sum_{k=1}^{n} \mathbb{E} \left( R \mathbb{1}_{[R > k]} \right). \]

The right side goes to zero as \( R \) has finite expectation.

(ii) By Kronecker’s lemma, it is enough to show that

\[ \sum_{n=1}^{\infty} \frac{1}{n} \mathbb{E} \left( \rho(R_n) \mathbb{1}_{[\rho(R_n) > n]} | F_{n-1} \right) < \infty \text{ almost surely.} \]

It is in turn enough to show that the above random variable has finite expectation. Now

\[ \sum_{n=1}^{\infty} \frac{1}{n} \mathbb{E} \left( \rho(R_n) \mathbb{1}_{[\rho(R_n) > n]} \right) \leq c \sum_{n=1}^{\infty} \frac{1}{n} \mathbb{E} \left( R \mathbb{1}_{[R > n]} \right) \]

using Assumption 2.1(iv-a)

\[ \leq c \sum_{n=1}^{\infty} \sum_{j=n}^{\infty} \frac{1}{n} \mathbb{E} \left( R \mathbb{1}_{[R \in (j, j+1]} \right) \]
\[
\leq c \sum_{j=1}^{\infty} \mathbb{E} \left( R \mathbb{1}_{[R \in (j, j+1)]} \right) \sum_{n=1}^{j} \frac{1}{n} \\
\leq c \sum_{j=1}^{\infty} \mathbb{E} \left( R \mathbb{1}_{[R \in (j, j+1)]} \right) (1 + \log j) \\
\leq c \sum_{j=1}^{\infty} \mathbb{E} \left( R \mathbb{1}_{[R \in (j, j+1)]} \right) (1 + \log_j R) \\
= c \mathbb{E} \left( R(1 + \log_j R) \right) < \infty.
\]

(iii) The result is obvious as each term is bounded by \( c' \frac{1}{n} \sum_{k=1}^{n} \mathbb{E} X \mathbb{1}_{[X > k]} \), which goes to zero.

(iv) Since \( \phi \) is eventually nondecreasing, we have, eventually,

\[
\mathbb{E} \left( \rho(R_k) \mathbb{1}_{[\rho(R_k) > k]} \big| \mathcal{F}_{k-1} \right) \leq \frac{1}{\phi(k)} \mathbb{E} \left( \rho(R_k) \phi(\rho(R_k)) \big| \mathcal{F}_{k-1} \right)
\]

and hence, we have

\[
\sum_{k=1}^{\infty} \frac{1}{k} \mathbb{E} \left( \rho(R_k) \mathbb{1}_{[\rho(R_k) > k]} \big| \mathcal{F}_{k-1} \right) \leq \sum_{k=1}^{\infty} \frac{1}{k\phi(k)} \mathbb{E} \left( \rho(R_k) \mathbb{1}_{[\rho(R_k) > k]} \big| \mathcal{F}_{k-1} \right).
\]

Since \( \sum_{n} 1/(n \phi(n)) < \infty \), under (2.2) the right side is finite; and under (2.3), the right side has finite expectation, and hence is finite almost surely. Finally, Kronecker's lemma gives us (2.6).

\[ \square \]

Remark 2.10. It is interesting to compare the assumptions in this article with those require to establish convergence in probability and \( L^1 \) convergence in Gangopadhyay and Maulik (2019). Note that Assumption 2.1(i)-(iii) are same as Assumption 3.1 of Gangopadhyay and Maulik (2019). As noted in Remark 2.2, the finite mean of the entries of the replacement matrices, as in Assumption 3.1 (ii) of Gangopadhyay and Maulik (2019) follows from Assumption 2.1(iv).

As noted in Remark 2.6, Assumption 2.1(iv) provides the analog of the uniform integrability of \( (\rho(R_n)) \) in Assumption 3.4 of Gangopadhyay and Maulik (2019). However, note that, in this article, we need to assume somewhat stronger conditions than uniform integrability, namely, either majorization
condition in Assumption 2.1(iv-a) or appropriate uniform moment bounds, as has been considered by Zhang (2012), in (2.3) of Assumption 2.1(iv-b).

Finally, Assumption 2.1(v) corresponds to Assumptions 3.2 and 3.3 of Gangopadhyay and Maulik (2019). While Assumption 3.2 of Gangopadhyay and Maulik (2019) on the properties of the matrix $H$ remain unchanged, we naturally strengthen the mode of convergence to almost sure from convergence in probability. We further require the truncated generating matrices $\tilde{H}_n$ to converge to $H$ instead of the generating matrices considered in Gangopadhyay and Maulik (2019). We have also seen in Remark 2.9, if we use convergence of $H_n$ as in Gangopadhyay and Maulik (2019), we need to additionally assume the negligibility of the tail conditional expectations in Cesaro sense, as given in (2.6). However, we have further proved in Proposition 2.1(iv) that, if we make Assumption 2.1(iv-b) as an analogue of the uniform integrability Assumption 3.4 of Gangopadhyay and Maulik (2019), then the convergence of the truncated generating matrices $\tilde{H}_n$ to $H$ is equivalent to that of the generating matrices $H_n$ themselves.

Remark 2.11. A model with i.i.d. $L\log L$-bounded replacement matrices was considered in Assumption 2.1(a) and the second part of Assumption 2.2(a) of Zhang (2012). Note that under i.i.d. setup, (2.5) on the generating matrices $H_n$ holds trivially with $H = E(R_1)$. In view of Remarks 2.3, 2.4, 2.9 and Proposition 2.1(i), the moment conditions can be substantially weakened to only the first moment condition. In fact the following set of assumptions are sufficient for Assumption 2.1, and hence will suffice to obtain Theorem 3.5 – see Corollary 3.6.

Assumption 2.2. The adapted sequence $((\chi_n, R_n), \mathcal{F}_n)$ has distribution, which satisfies:

(i) The initial configuration $C_0$ is nonzero with nonnegative entries and finite mean.

(ii) The replacement matrices $R_n$ are i.i.d. with $E(\rho(R_1)) < \infty$.

(iii) The colors chosen form the sequence $(\chi_n)$, obtained using an i.i.d. Uniform$(0, 1)$ sequence $(U_n)$ independent of the sequence $(R_n)$ as:

$$\chi_n = \sum_{i=1}^{d} e_i 1_{U_n \in \{ \sum_{j=1}^{\ell} C_{n-1,j}/S_{n-1}, \sum_{j=1}^{\ell} C_{n-1,j}/S_{n-1} \}}.$$

Note that the setup in Assumption 2.2 was considered in Chapter V.9.3 of Athreya and Ney (1972), and again in Theorem 2.1(a) of Zhang (2012), but with additional $L\log L$ moment assumptions in Assumption 2.2(ii). The above set of assumptions achieves a significant weakening of that.
2.2. **Useful Results.** This subsection collects miscellaneous results which we shall use in the later subsections. We start with couple of properties of $\rho(R_n)$, which follow from Assumption 2.1(iv).

**Lemma 2.2.** Under Assumption 2.1(iv), $\rho(R_n)$ satisfy:

(i) $\mathbb{P}(\rho(R_n) > n \text{ infinitely often } n) = 0$ and hence $\mathbb{1}_{[\rho(R_n) > n]} = 0$ eventually in $n$ almost surely.

(ii) $\sum_{n=1}^{\infty} \frac{1}{n^2} \mathbb{E}\left(\rho(R_n)^2 \mathbb{1}_{[\rho(R_n) \leq n]} \bigg| F_{n-1}\right) < \infty$ almost surely.

**Proof.** (i) Under Assumption 2.1(iv-a), we use the majorization inequality and finiteness of expectation of $R$. Then the result follows using the first Borel-Cantelli lemma.

Under Assumption 2.1(iv-b), we note that $1/(n\phi(n))$ is summable and, using Markov type inequality, that

$$
(2.7) \quad \mathbb{P}(\rho(R_n) > n|F_{n-1}) \leq \frac{1}{n\phi(n)} \mathbb{E}(\rho(R_n)\phi(\rho(R_n))|F_{n-1}).
$$

When the conditional expectations are bounded, the result follows by the conditional version of the first Borel-Cantelli lemma. When the unconditional expectations are bounded, we take further expectation of (2.7) and apply the first Borel-Cantelli lemma.

(ii) First note that it is enough to show that the expected value of the random variable is finite. Next, observe that, under Assumption 2.1(iv-a), we have

$$
\sum_{n \geq 1} \frac{1}{n^2} \mathbb{E}\left(\rho(R_n)^2 \mathbb{1}_{[\rho(R_n) \leq n]} \right) = \sum_{n \geq 1} \frac{1}{n^2} \int_0^n 2r \mathbb{P}(\rho(R_n) > r)dr
$$

$$
= \sum_{n \geq 1} \frac{1}{n^2} \sum_{j=1}^{\lfloor n \rfloor} \int_{j-1}^j 2r \mathbb{P}(\rho(R_n) > r)dr
$$

$$
\leq c \sum_{n \geq 1} \frac{1}{n^2} \sum_{j=1}^{\lfloor n \rfloor} \int_{j-1}^j 2r \mathbb{P}(R > r)dr
$$

$$
= c \sum_{j \geq 1} \sum_{n \geq j} \frac{1}{n^2} \int_{j-1}^j 2r \mathbb{P}(R > r)dr
$$

$$
\leq 4c \sum_{j \geq 1} \frac{1}{j} \int_{j-1}^j 2r \mathbb{P}(R > r)dr
$$

$$
\leq 8c \sum_{j \geq 1} \int_{j-1}^j \mathbb{P}(R > r)dr
$$
CONVERGENCE OF RANDOMIZED URN MODELS

\[ = 8c \int_0^\infty \mathbb{P}(R > r)dr < \infty. \]

Under Assumption 2.1(iv-b), for all large enough \( n \), we have

\[ \mathbb{E} \left( \rho(R_n)^2 \mathbb{1}_{[\rho(R_n) \leq n]} \big| \mathcal{F}_{n-1} \right) \leq \frac{n}{\phi(n)} \mathbb{E} \left( \rho(R_n) \phi(\rho(R_n)) \big| \mathcal{F}_{n-1} \right), \]

when \( x/\phi(x) \) is eventually monotone nondecreasing. Also, the left side of (2.8) is bounded by a constant multiple of \( \mathbb{E} \left( \rho(R_n) \phi(\rho(R_n)) \big| \mathcal{F}_{n-1} \right) \) when \( x/\phi(x) \) is bounded. In either case, when the conditional expectations are bounded in Assumption 2.1(iv-b), the required result follows as both \( 1/\left( n\phi(n) \right) \) and \( 1/n^2 \) are summable. When the expectations are bounded in Assumption 2.1(iv-b), we take further expectation on the bounds and the result follows from summability of \( 1/\left( n\phi(n) \right) \) and \( 1/n^2 \).

\[ \square \]

The next result provides kind of a martingale convergence theorem where the terms are centered by a truncated conditional expectation.

**Proposition 2.3.** Let \((X_n, \mathcal{F}_n)\) be a adapted sequence, such that \(|X_n| \leq c'' \rho(R_n)\) for some \( c'' \in (0, \infty) \). Then, under Assumption 2.1(iv), with probability 1, we have

\[ \sum_{k=1}^{n} \frac{1}{k} \left( X_k - \mathbb{E} \left( X_k \mathbb{1}_{[\rho(R_k) \leq k]} \big| \mathcal{F}_{k-1} \right) \right) \text{ converges, and} \]

\[ \frac{1}{n} \sum_{k=1}^{n} \left( X_k - \mathbb{E} \left( X_k \mathbb{1}_{[\rho(R_k) \leq k]} \big| \mathcal{F}_{k-1} \right) \right) \to 0. \]

**Proof.** We write

\[ \sum_{k=1}^{n} \frac{1}{k} \left( X_k - \mathbb{E} \left( X_k \mathbb{1}_{[\rho(R_k) \leq k]} \big| \mathcal{F}_{k-1} \right) \right) \]

\[ = \sum_{k=1}^{n} \frac{1}{k} \left( X_k \mathbb{1}_{[\rho(R_k) \leq k]} - \mathbb{E} \left( X_k \mathbb{1}_{[\rho(R_k) \leq k]} \big| \mathcal{F}_{k-1} \right) \right) + \sum_{k=1}^{n} \frac{1}{k} X_k \mathbb{1}_{[\rho(R_k) > k]} \]

All terms of the second sum vanish eventually almost surely by Lemma 2.2(i) and hence will converge. Observe the first sum is a martingale and hence to show its almost sure convergence, it is enough to check that, almost surely,

\[ \sum_{n} \frac{1}{n^2} \mathbb{E} \left( X_n^2 \mathbb{1}_{[\rho(R_n) \leq n]} \big| \mathcal{F}_{n-1} \right) < \infty. \]
This follows immediately from Lemma 2.2(ii), since $|X_n| \leq c'' \rho (R_n)$. \qed

2.3. Stochastic Approximation. Stochastic approximation developed as a tool from the works of Robbins and Monro (1951) and Kiefer and Wolfowitz (1952). Stochastic approximation has been mostly applied using either of two methods, namely, the ODE method considered in, for example, Kushner and Clark (1978) or Duflo (1997), and the dynamical system method considered in, for example, Benaîm (1999) or Borkar (2008). We shall be using the former one. In this subsection, we begin by stating the main result of Stochastic Approximation which we use in this article. The result is essentially a special case of Theorem 9.2.8 of Duflo (1997), which in turn was adapted from Kushner and Clark (1978). Note that the result is about deterministic sequences. This turns out to be enough for our application because we are only concerned with almost sure convergence.

**Theorem 2.4.** Let $(x_n)_{n=0}^\infty$ be a sequence that takes values in a compact convex subset $S$ of $\mathbb{R}^K$. Let $(a_n)_{n=0}^\infty$ be a sequence of positive real numbers, called the step sizes, such that $\lim_{n \to \infty} a_n = 0$ and $\sum_{n=0}^\infty a_n = \infty$. Define for $n \geq 0$ and $T > 0$

$$\tau_n(T) := \inf \{ k \geq n : a_n + \cdots + a_k \geq T \}.$$  

Let $(\gamma_n)_{n=0}^\infty$ be a sequence, called the error sequence, in $\mathbb{R}^K$ such that for all $T > 0$

$$\lim_{n \to \infty} \sup_{n \leq j \leq \tau_n(T)} \left\| \sum_{k=n}^j a_k \gamma_k \right\|_{\ell^1} = 0.$$  

Let $h : S \to \mathbb{R}^K$ be a continuous function. Suppose for all $n \geq 0$

$$x_{n+1} = x_n + a_n h(x_n) + a_n \gamma_n.$$  

If the ODE $\dot{x} = h(x)$ has a unique solution $z : \mathbb{R} \to S$, then $\lim_{n \to \infty} x_n = z(0)$.

**Remark 2.12.** As noted in Duflo (1997), (2.10) holds if in particular $\sum_n a_n \gamma_n$ is convergent. Note that it also holds if $\gamma_n \to 0$ because

$$\sup_{n \leq j \leq \tau_n(T)} \left\| \sum_{k=n}^j a_k \gamma_k \right\|_{\ell^1} \leq \sum_{j=n}^{\tau_n(T)} a_j \sup_{j \geq n} \left\| \gamma_j \right\|_{\ell^1} \leq (T + a_{\tau_n(T)}) \sup_{j \geq n} \left\| \gamma_j \right\|_{\ell^1}.$$
Note that, if \( a_n \) is monotone, then the convergence of \( \sum a_n \gamma_n \) implies, by Kronecker lemma, that \( a_n \Gamma_n \to 0 \), where \( \Gamma_n := \sum_{i=1}^{n} \gamma_i \). Further, if \( a_n \) is of the order \( 1/n \), then \( \gamma_n \to 0 \) implies \( a_n \Gamma_n \to 0 \) as well. In fact, we can replace the conditions in Remark 2.12 by the weaker \( a_n \Gamma_n \to 0 \) in certain cases.

**Proposition 2.5.** If the step size sequence \((a_n)\) of a stochastic approximation is nonincreasing and the error sequence \((\gamma_n)\) satisfies \( a_n \Gamma_n \to 0 \), then, for any \( T > 0 \),

\[
\sup_{n \leq j \leq \tau_n(T)} \left\| \sum_{k=n}^j a_k \gamma_k \right\|_{\ell^1} \leq 2 + \frac{a_n}{a_{\tau_n(T)}}.
\]

**Proof.** Fix \( T > 0 \). First observe that

\[
a_n \gamma_n = (a_n \Gamma_n - a_{n-1} \Gamma_{n-1}) + a_n a_{n-1} \Gamma_{n-1} \left( \frac{1}{a_n} - \frac{1}{a_{n-1}} \right),
\]

which leads to, for \( n \leq j \leq \tau_n(T) \), using the fact that \((a_n)\) is nonincreasing,

\[
\left\| \sum_{k=n}^j a_k \gamma_k \right\|_{\ell^1} \leq \left\| \sum_{k=n}^j (a_k \Gamma_k - a_{k-1} \Gamma_{k-1}) \right\|_{\ell^1} + a_n \sup_{k \geq n-1} \| a_k \Gamma_k \|_{\ell^1} \sum_{k=n}^j \left( \frac{1}{a_k} - \frac{1}{a_{k-1}} \right)
\]

\[
\leq \| a_j \Gamma_j \|_{\ell^1} + \| a_{n-1} \Gamma_{n-1} \|_{\ell^1} + \sup_{k \geq n-1} \| a_k \Gamma_k \|_{\ell^1} \frac{a_n}{a_j}.
\]

Hence, we have

\[
\sup_{n \leq j \leq \tau_n(T)} \left\| \sum_{k=n}^j a_k \gamma_k \right\|_{\ell^1} \leq \sup_{j \geq n-1} \| a_j \Gamma_j \|_{\ell^1} \left( 2 + \frac{a_n}{a_{\tau_n(T)}} \right)
\]

and the result follows as \( a_n \Gamma_n \to 0 \). \( \square \)

Further simplification is possible and (2.10) will hold, if additionally \( a_n \) is of order \( 1/n \).

**Corollary 2.6.** If the step size sequence \((a_n)\) of a stochastic approximation is nonincreasing,

\[
0 < \lim \inf \, n a_n \leq \lim \sup \, n a_n < \infty
\]

(2.11)
and the error sequence \((\gamma_n)\) satisfies \(\frac{1}{n} \Gamma_n \to 0\), then

\[
\limsup_{n \to \infty} \sup_{n \leq j \leq \tau_n(T)} \left\| \sum_{k=n}^{j} a_k \gamma_k \right\|_{\ell^1} = 0.
\]

**Proof.** First observe that (2.11) implies \(a_n \Gamma_n \to 0\). Also note that

\[
a_n \tau_n(T) = n a_n \cdot \frac{\tau_n(T)}{n} \cdot \frac{1}{\tau_n(T) a_{\tau_n(T)}}.
\]

Thus, by (2.11) and Proposition 2.5, it is enough to show that \(\tau_n(T)/n\) is bounded for any \(T > 0\). Since \(\liminf n a_n > 0\), choose \(a^*\) such that \(n a_n > a^*\) for all large enough \(n\). By definition of \(\tau_n(T)\) in (2.9), observe that

\[
a^* + \ldots + \frac{a^*}{\tau_n(T)} \leq a_n + \ldots + a_{\tau_n(T)} \leq T,
\]

which gives \(\log \tau_n(T) - \log(n - 1) - 1 \leq T/a^*\), or equivalently, \(\tau_n(T)/n \leq \tau_n(T)/(n - 1) \leq \exp(T/a^* + 1)\), as required.

We shall actually check the conditions of Corollary 2.6 to apply Theorem 2.4.

3. Analysis of Urn Model. In this section, we rewrite the evolution equation (2.1) in a form appropriate for stochastic approximation, while the following two subsections analyze the step size and the error terms respectively. Finally, we provide the main theorem on almost sure and \(L^1\) convergence for urn models.

3.1. Stochastic Approximation Equation. In this subsection, we write the evolution equation (2.1) appropriately for stochastic approximation. For that, we need to study the total amount added to the urn in each step, denoted by \((Y_n)_{n=0}^\infty\). Define \(Y_0 := S_0\) and for \(n \geq 1\), \(Y_n := S_n - S_{n-1} = \chi_n R_n 1^T\). It is immediate to observe that

\[
0 \leq Y_n = \chi_n R_n 1^T \leq \rho(R_n) \text{ almost surely.}
\]

Following Gangopadhyay and Maulik (2019), we rewrite the evolution equation (2.1) as

\[
\frac{C_n}{S_n} = \frac{C_{n-1}}{S_{n-1}} + \frac{1}{S_n} h_H \left( \frac{C_{n-1}}{S_{n-1}} \right) + \frac{\delta_n}{S_n} + \frac{\xi_n}{S_n},
\]
where \( h_H, \delta_n, \xi_n \) are defined as follows. The drift \( h_H \), indexed by \( K \times K \) matrices, is defined as

\[
h_H(X) := XH - X \left( XH^T \right).
\]

For each \( n \geq 1 \), define the martingale difference term

\[
\delta_n := \left( \chi_n R_n - \frac{C_{n-1}}{S_{n-1}} Y_n \right) - \mathbb{E} \left( \left( \chi_n R_n - \frac{C_{n-1}}{S_{n-1}} Y_n \right) 1_{|\rho(R_n)| \leq n} \mid \mathcal{F}_{n-1} \right),
\]

and adjusted truncated conditional expectation term

\[
\xi_n := \frac{C_{n-1}}{S_{n-1}} \left( \tilde{H}_{n-1} - H \right) - \frac{C_{n-1}}{S_{n-1}} \left( \frac{C_{n-1}}{S_{n-1}} \left( \tilde{H}_{n-1} - H \right) 1^T \right) = h_{H_{n-1}} \left( \frac{C_{n-1}}{S_{n-1}} \right),
\]

We shall refer to (3.2) as the stochastic approximation equation. Note that the stochastic approximation for the sequence \( (C_n/S_n) \) takes values in the closed bounded convex set of probability simplex in \( \mathbb{R}^K \). The corresponding ODE defined in terms of \( h_H \) is a first order quadratic equation of Lotka-Volterra type. We quote its solution from Gangopadhyay and Maulik (2019).

**Theorem 3.1** (Gangopadhyay and Maulik (2019), Proposition 3.3). The only solution of \( \dot{x} = h_H(x) \), where \( x(t) \) is a probability vector for all \( t \), is \( x(t) = \pi_H \) for all \( t \).

Also, since \( h_H(x) \) is polynomial in \( x \), it is continuous, as required in Theorem 2.4.

3.2. **Step Size.** Unlike the usual step size \( 1/n \), we use, as in Gangopadhyay and Maulik (2019) the random step size \( 1/S_{n+1} \) of the stochastic approximation equation (3.2). The step size \( 1/S_{n+1} \) has also been used to analyze two color urn models with fixed replacement unbalanced matrices in Renlund (2010, 2011). The choice of this random step size allows us to consider \( C_n/S_n \) as the stochastic approximation sequence. Note that \( C_n/S_n \) being a proportion vector takes values in the \( k \)-dimensional probability simplex, which is compact and convex. Since \( S_n \)’s are partial sum of nonnegative \( Y_n \)’s, they are nondecreasing and hence the step sizes are nonincreasing, as required in Corollary 2.6. In the next proposition, we check the remaining properties of the step size, including the rate of decay as required in Corollary 2.6.
Proposition 3.2. Under Assumption 2.1, the following results hold almost surely for $S_n$:

(i) $S_n/n - \frac{1}{n} \sum_{m=1}^{n} \mathbb{E} (Y_m 1_{[\rho(R_m) \leq m]} | \mathcal{F}_{m-1}) \to 0$;

(ii) $0 < \sigma(H) \leq \lim inf_{n \to \infty} \frac{S_n}{n} \leq \lim sup_{n \to \infty} \frac{S_n}{n} \leq \rho(H) < \infty$;

(iii) $S_n \to \infty$;

(iv) $\sum_{n} \frac{1}{S_n} = \infty$.

Proof. Since $0 \leq Y_n \leq \rho(R_n)$, using Proposition 2.3, (i) above follows directly. Define

$$\eta_n := \frac{S_n}{n} - \frac{1}{n} \sum_{m=1}^{n} \mathbb{E} (Y_m 1_{[\rho(R_m) \leq m]} | \mathcal{F}_{m-1}) + \frac{1}{n} \sum_{m=1}^{n} \frac{C_{m-1}}{S_{m-1}} (\tilde{H}_{m-1} - H) 1^T.$$ 

From (i), we have that the sum of the first two terms converges to zero almost surely. The third term is bounded by $\frac{1}{n} \sum_{m=1}^{n} \rho (\tilde{H}_{m-1} - H)$, which converges to zero almost surely by Assumption 2.1(v). Thus $\eta_n \to 0$ almost surely. Further, we have

$$0 < \sigma(H) \leq \left| \frac{S_n}{n} - \eta_n \right| \leq \rho(H) < \infty,$$

which proves (ii). Rest of the result then follow immediately.

3.3. Error terms. In this subsection, we show Cesaro negligibility of the error terms ($\delta_n$) and ($\xi_n$) of the stochastic approximation equation (3.2), as required in Corollary 2.6.

We next check the Cesaro negligibility of ($\delta_n$).

Lemma 3.3. Under Assumption 2.1, we have $\frac{1}{n} \sum_{m=1}^{n} \delta_m \to 0$ almost surely.

Proof. It is easy to note that each coordinate of $(\chi_n R_n - C_{n-1}/S_{n-1} Y_n)$ is bounded by $2\rho(R_n)$. Hence, using Proposition 2.3, the negligibility condition holds for each coordinate of $\frac{1}{n} \sum_{m=1}^{n} \delta_m$.

We next check the Cesaro negligibility of ($\xi_n$).

Lemma 3.4. Under Assumption 2.1, we have $\frac{1}{n} \sum_{m=1}^{n} \xi_m \to 0$ almost surely.
Proof. Recall that
\[
\xi_n = C_n \left( H_{n-1} - H \right) - S_n \left( \frac{C_n}{S_n} \left( H_{n-1} - H \right)^T \right).
\]
Then \( \|\xi_n\|_{\ell^1} \leq 2\rho \left( H_{n-1} - H \right) \), since \( C_n/S_n \) is a probability vector. By Assumption 2.1(v), the result follows. \( \square \)

3.4. Main Results. Finally, in this Subsection, we prove almost sure and \( L^1 \) convergence for urn model.

**Theorem 3.5.** Under Assumption 2.1, the following almost sure convergence happen:

1. \( \frac{C_n}{S_n} \to \pi_H \);
2. \( \frac{S_n}{n} \to \lambda_H \);
3. \( \frac{C_n}{n} \to \lambda_H \pi_H \);
4. \( \frac{N_n}{n} \to \pi_H \).

The convergence given above are also in \( L^1 \), if one of the uniform integrability assumptions, namely Assumption 2.1(iv-a) or condition (2.3) of Assumption 2.1(iv-b), holds.

Proof. We have already noted that the ODE associated with stochastic approximation (3.2) has unique probability solution \( \pi_H \) in Theorem 3.1. The conditions on step sizes and the error terms of the stochastic approximation equation (3.2) has been checked in Subsections 3.2 and 3.3. Hence the almost sure convergence of the proportion vector to \( \pi_H \) in (3.4) holds, using Theorem 2.4 and Corollary 2.6.

Next we consider (3.5). Using Proposition 3.2(i), it is enough to show that

\[
\frac{1}{n} \sum_{m=1}^{n} \mathbb{E} \left( Y_m \mathbb{1}[\rho(R_m) \leq m] \big| \mathcal{F}_{m-1} \right) \to \lambda_H
\]

almost surely. Recall that

\[
\frac{1}{n} \sum_{m=1}^{n} \mathbb{E} \left( Y_m \mathbb{1}[\rho(R_m) \leq m] \big| \mathcal{F}_{m-1} \right)
\]
\[
= \frac{1}{n} \sum_{m=1}^{n} \frac{C_{m-1}}{S_{m-1}} (\bar{H}_{m-1} - H) \mathbf{1}^{T} + \frac{1}{n} \sum_{m=1}^{n} \frac{C_{m-1}}{S_{m-1}} H \mathbf{1}^{T}.
\]

The first term on the right side converges to 0 almost surely by Assumption 2.1(v). Further, since \(C_n/S_n \to \pi_H\) almost surely, we get
\[
\frac{C_n}{S_n} H \mathbf{1}^{T} \to \pi_H H \mathbf{1}^{T} = \lambda_H
\]
almost surely and (3.8) follows using Cesaro. Then, (3.6) follows easily from (3.4) and (3.5).

Using Cesaro limit of the convergence in (3.4), 
\[
\frac{1}{n} \sum_{m=0}^{n-1} \frac{C_m}{S_m} \to \pi_H\]
almost surely. Finally, to prove (3.7), we shall show that, almost surely
\[
\frac{N_n}{n} - \frac{1}{n} \sum_{m=0}^{n-1} \frac{C_m}{S_m} = \frac{1}{n} \sum_{m=1}^{n} \left( X_m - \frac{C_{m-1}}{S_{m-1}} \right) \to 0.
\]

By Kronecker’s Lemma, it is enough to show that
\[
\sum_{n} \frac{1}{n} \left( X_n - \frac{C_{n-1}}{S_{n-1}} \right)
\]
is convergent almost surely, which holds since \(\|X_n/n\|_1^2 = 1/n^2\) is summable.

Next we consider \(L^1\) convergence. Since \(C_n/S_n\) and \(N_n/n\) are bounded, the convergences in (3.4) and (3.7) hold in \(L^1\) as well by Dominated Convergence Theorem. If the convergence in (3.5) holds in \(L^1\), then by Lemma 3.5 of Gangopadhyay and Maulik (2019), the convergence in (3.6) holds as well in \(L^1\). For (3.5), observe that, from (3.1), \(0 \leq Y_n \leq \rho(R_n)\) for \(n \geq 1\) and \(Y_0 = C_0 \mathbf{1}^{T}\) is integrable by Assumption 2.1(i). Further, as noted in Remark 2.6, Assumption 2.1(iv-a) or condition (2.3) of Assumption 2.1(iv-b) implies uniform integrability of \((\rho(R_n))\). Hence \((Y_n)_{n=0}^{\infty},\) and thus \(S_n/n = \sum_{k=0}^{n} Y_k/n\) are also uniformly integrable. So the almost sure convergence in (3.5) becomes \(L^1\) as well.

Theorem 3.5 holds under further sufficient conditions as has been noted earlier. In particular, as has been noted in Remark 2.11 it holds under i.i.d. setup with only first moment condition. This is a significant improvement of the almost sure convergence result obtained by Athreya and Ney (1972), and later by Zhang (2012), for i.i.d. replacement matrices under \(L \log_n L\) moment condition.

**Corollary 3.6.** Theorem 3.5 holds under Assumption 2.2. The convergence is both almost sure and in \(L^1\).
In fact, as has been seen in Remark 2.9 and Proposition 2.1(ii), $L \log_+ L$ condition satisfies even under a more general adaptive setup and the generating matrices themselves converge.

**Corollary 3.7.** Assumption 2.1(i)–(iii), (iv-a) with $E(R \log_+ R) < \infty$ and the convergence condition (2.5) imply Theorem 3.5. The convergence is both almost sure and in $L^1$.

Zhang (2012), under his Assumption 2.1 (a) and the second part of Assumption 2.2. (b) considered an adaptive model with bounded $L(\log_+ L)^p$ moment condition. In Remark 2.9 and Proposition 2.1(iv), we have seen a similar and more general moment condition, namely Assumption 2.1(iv-b), suffices.

**Corollary 3.8.** Assumption 2.1(i)–(iii), (iv-b) and the convergence condition (2.5) imply Theorem 3.5. The convergence is almost sure. It is further in $L^1$ if condition (2.3) of Assumption 2.1(iv-b) holds.

**References.**

Athreya, K. B. and Ney, P. E. (1972). *Branching processes*. Springer-Verlag, New York-Heidelberg. Die Grundlehren der mathematischen Wissenschaften, Band 196. MR0373940

Bai, Z. D. and Hu, F. (2005). Asymptotics in randomized URN models. *Ann. Appl. Probab.* 15 914–940. MR2114994

Benaïm, M. (1999). Dynamics of stochastic approximation algorithms. In *Séminaire de Probabilités, XXXIII*. Lecture Notes in Math. 1709 1–68. Springer, Berlin. MR1767993

Borkar, V. S. (2008). *Stochastic approximation*. Cambridge University Press, Cambridge; Hindustan Book Agency, New Delhi A dynamical systems viewpoint. MR2442439

Duflo, M. (1997). *Random iterative models. Applications of Mathematics (New York).* 34. Springer-Verlag, Berlin Translated from the 1990 French original by Stephen S. Wilson and revised by the author. MR1485774

Gangopadhyay, U. and Maulik, K. (2019). Stochastic approximation with random step sizes and urn models with random replacement matrices having finite mean. *Ann. Appl. Probab.* 29 2033–2066.

Kiefer, J. and Wolfowitz, J. (1952). Stochastic estimation of the maximum of a regression function. *Ann. Math. Statistics* 23 462–466. MR050243

Kushner, H. J. and Clark, D. S. (1978). *Stochastic approximation methods for constrained and unconstrained systems*. Applied Mathematical Sciences 26. Springer-Verlag, New York-Berlin. MR499560

Laruelle, S. and Pagès, G. (2013). Randomized urn models revisited using stochastic approximation. *Ann. Appl. Probab.* 23 1409–1436. MR3098437

Pólya, G. (1930). Sur quelques points de la théorie des probabilités. *Ann. Inst. H. Poincaré* 1 117–161. MR1507985

Pólya, G. and Eggenberger, F. (1923). Über die Statistik verketteter Vorgänge. Z. Angew. Math. Mech. 3 279–289.

Renlund, H. (2010). Generalized Polya urns via stochastic approximation. arxiv:1002.3716.
RENlund, H. (2011). Limit theorems for stochastic approximation algorithms. arxiv:1102.4741.
Robbins, H. and Monro, S. (1951). A stochastic approximation method. Ann. Math. Statistics 22 400–407. MR0042668
Zhang, L. (2012). The Gaussian approximation for generalized Friedman’s urn model with heterogeneous and unbalanced updating. Sci. China Math. 55 2379–2404. MR2994126
Zhang, L. X. (2016). Central limit theorems of a recursive stochastic algorithm with applications to adaptive designs. Ann. Appl. Probab. 26 3630–3658. MR3582813

UJAN Gangopadhyay
Department of Mathematics
University of Southern California
Los Angeles, CA
USA
E-mail: ujangangopadhyay@gmail.com

Krishanu Maulik
Theoretical Statistics and Mathematics Unit
Indian Statistical Institute
Kolkata
India
E-mail: krishanu@isical.ac.in
URL: http://www.isical.ac.in/~krishanu