Submitted to the Annals of Applied Probability

A NEW APPROACH TO PÓLYA URN SCHEMES AND ITS INFINITE COLOR GENERALIZATION

BY ANTAR BANDYOPADHYAY[1], DEBLEENA THACKER[2,*]

[1]*Theoretical Statistics and Mathematics Unit*
Indian Statistical Institute, Kolkata and Delhi Centre
7 S. J. S. Sansanwal Marg
New Delhi 110016
INDIA, antar@isid.ac.in

[2]*Department of Mathematical Sciences*
Durham University
Mathematical Sciences and Computer Science Building MCS2030
Upper Mountjoy Campus, Stockton Road
Durham DH1 3LE
*UNITED KINGDOM, *thackerdebleena@gmail.com*

In this work, we introduce a generalization of the classical Pólya urn scheme [40] with colors indexed by a Polish space, say, S. The urns are defined as finite measures on S endowed with the Borel σ-algebra, say, $\mathcal{S}$. The generalization is an extension of a model introduced earlier by Blackwell and MacQueen [8]. We present a novel approach of representing the observed sequence of colors from such a scheme in terms an associated *branching Markov chain* on the *random recursive tree*. The work presents fairly general asymptotic results for this new generalized urn models. As special cases we show that the results on classical urns, as well as, some of the results proved recently for infinite color urn models in [6, 5], can easily be derived using the general asymptotic. We also demonstrate some newer results for infinite color urns.

CONTENTS

MSC2020 subject classifications: Primary 60F05, 60F10; secondary 60G50.
Keywords and phrases: branching Markov chains, color count statistics, infinite color urn, random recursive trees, random replacement matrices, reinforcement processes, representation theorem, urn models.

1. Introduction. In recent days various *urn schemes* and their many generalizations have been a key element of study for random processes with reinforcements [29, 3, 39, 23, 10, 18, 16, 15, 33, 13, 6, 5]. Starting from the seminal work by Pólya [40], various types of urn schemes with finitely many colors have been widely studied in literature [25, 24, 1, 2, 39, 38, 27, 28, 29, 3, 23, 10, 11, 18, 14, 13, 34]. [39] and [34] provide extensive surveys of the known results. However, other than the classical work by Blackwell and MacQueen [8], there has not been much development of infinite color generalization of the Pólya urn scheme. Recently the authors studied a specific class of urn models with infinitely many colors where the color set is indexed by the d-dimensional integer lattice $\mathbb{Z}^d$, [6, 5]. These works nicely complement the work [8] by introducing examples of infinite color schemes with "*off-diagonal*" entries and showed that the asymptotic behavior is essentially determined by an underlying random walk.

In this paper, we further generalize urn schemes with colors index by an arbitrary set S endowed with a σ-algebra $\mathcal{S}$. As we will see in the sequel, the classical models can be realized as a sub-model when S is finite and in that case $\mathcal{S}$ will simply be the power set of S, which we will denote by $\wp(S)$. The non-classical case discussed in [8] can also be obtained by appropriately choosing the measurable space $(S, \mathcal{S})$ as the *Borel space* of a Polish space S. Further the models described in [6, 5] can be obtained by choosing $S = \mathbb{Z}^d$ and $\mathcal{S} = \wp\left(\mathbb{Z}^d\right)$.

We will only consider *balanced* urn schemes. For S countable (finite or infinite), it means that if $R := ((R(i,j)))_{i,j \in S}$ denotes the *replacement matrix*, that is, $R(i,j) \geq 0$ is the *number of balls of color j to be placed in the the urn when the color of the selected ball is i*, then for a balanced urn, all row sums of R are constant. In this case, without loss of generality, it is somewhat customary to assume that R is a *stochastic matrix* [10, 11]. For more general S we refer to the next subsection for further details.

1.1. *Model.* We consider the following generalization of Pólya urn scheme where the colors are indexed by a non-empty subset S of $\mathbb{R}^d$ for some $d \geq 1$, such that, under subspace topology S is a Polish space. A necessary and sufficient condition for S to be Polish is that it is a $G_\delta-$set, that is, S is a countable intersection of open sets, [12]. We endow S with the corresponding Borel σ-algebra and denote it by $\mathcal{S}$. Let $\mathcal{M}(S)$ and $\mathcal{P}(S)$ denote respectively the set of all *finite measures* and the set of all *probability measures* on the measurable space $(S,\mathcal{S})$. Note that the classical case when S is finite or the non-classical cases discussed in [6, 5] are obtained by taking S as a discrete subset of $\mathbb{R}^d$ of appropriate cardinality.

We would like to note here that many of our results goes through for S a general state space endowed with a σ-algebra $\mathcal{S}$. The assumptions of S is a Polish space and/or a subset of an Euclidean space are then not needed. For more details we refer the readers to our Remark 3.3 in Section 3.

Let $R : S \times \mathcal{S} \to [0,1]$ be a Markov kernel on S, that is, for every $s \in S$, as a set-function of $\mathcal{S}$, $R(s,\cdot)$ is a probability measure on $(S,\mathcal{S})$; and for every $A \in \mathcal{S}$, the function $s \mapsto R(s,A)$ is $\mathcal{S}/\mathcal{B}_{[0,1]}$-measurable. Our main objective will be to study the following three quantities:

(I) Random Configuration of the Urn: The random configuration of the urn at time $n \geq 0$ is a random finite measure $U_n \in \mathcal{M}(S)$, with total mass as $n+1$, such that, if Z_n represents the randomly chosen color at the $(n+1)$-th draw then the conditional distribution of Z_n given the "*past*", is given by

$$\mathbf{P}\left(Z_n \in \cdot \,\middle|\, U_n, U_{n-1}, \cdots, U_0\right) \propto U_n(\cdot)\,.$$

Formally, starting with $U_0 \in \mathcal{P}(S)$, we define $(U_n)_{n\geq 0} \subseteq \mathcal{M}(S)$ recursively as follows

$$U_{n+1}(A) = U_n(A) + R(Z_n, A)\,, \qquad A \in \mathcal{S}, \tag{1.1}$$

and

$$\mathbf{P}\left(Z_n \in \cdot \,\middle|\, U_n, U_{n-1}, \cdots, U_0\right) = \frac{U_n(\cdot)}{n+1}. \tag{1.2}$$

Notice that, if S is countable then R can be presented as a *stochastic matrix* and then $R(Z_n,\cdot)$ is the Z_n-th row of the replacement matrix R. We will refer to the process $(U_n)_{n\geq 0}$ as the *urn model* with colors index by S, initial configuration U_0 and replacement kernel R.

Observe that, one can associate with every urn model a Markov chain $(X_n)_{n\geq 0}$ on the state space S, with transition kernel R and initial distribution U_0. Moreover, without loss we can assume that the our underlying probability space is large enough, to have this chain independent of the urn process $(U_n)_{n\geq 0}$. Conversely, given any Markov chain $(X_n)_{n\geq 0}$, on the state space S, with transition kernel R and an initial distribution U_0, one can associate a balanced urn model $(U_n)_{n\geq 0}$, which will be independent of the chain. We will call this Markov chain $(X_n)_{n\geq 0}$ as the Markov chain associated with the urn model $(U_n)_{n\geq 0}$.

It is worth mentioning here that a little more general model may be obtained by taking $U_0 \in \mathcal{M}(S)$ and not just in $\mathcal{P}(S)$. However, asymptotic results for U_n, when $U_0 \in \mathcal{M}(S)$ can be easily derived from the special case $U_0 \in \mathcal{P}(S)$.

(II) Expected Configuration of the Urn: From equation (1.2) and taking expectation, we get

$$\mathbf{E}[U_n](S) = \mathbf{E}[U_n(S)] = n+1, \tag{1.3}$$

for all $n \geq 0$. Thus $\frac{\mathbf{E}[U_n]}{n+1}$ is also a probability measure on $(S,\mathcal{S})$. In fact, it is the distribution of Z_n, the $(n+1)$-th selected color. This follows by taking expectation on both sides of equation (1.2),

$$\mathbf{P}(Z_n \in A) = \frac{\mathbf{E}[U_n(A)]}{n+1}, \;\; A \in \mathcal{S}. \tag{1.4}$$

(III) Color count statistics: The *Color count statistics* is defined as follows:

$$\mathcal{N}_n := \sum_{k=0}^{n-1} \delta_{Z_k}, \tag{1.5}$$

where δ_z stands for the *Dirac delta* measure on $(S, \mathcal{S})$, which is degenerated at $z \in S$. Naturally, $\mathcal{N}_n$ is the *empirical measure* of the observed color sequence $(Z_k)_{k=0}^{n-1}$. In the classical setup when S is a finite set, $\mathcal{N}_n$ can be viewed as the frequency vector of the observed sequence of colors upto time $n-1$. This quantity has also been studied in the classical urn model literature [1, 2, 3]. In our general setup $\mathcal{N}_n$ is a random *counting measure* defined on $(S, \mathcal{S})$ with $\mathcal{N}_n(S) = n$. So naturally, $\frac{\mathcal{N}_n}{n}$ is a random probability measure on $(S, \mathcal{S})$. It follows from (1.1)

$$\frac{U_n}{n+1} = \frac{U_0}{n+1} + \frac{1}{n+1} \mathcal{N}_n R. \tag{1.6}$$

Note that we use the convention that for a finite measure $\mu \in \mathcal{M}(S)$, the expression $\mu R(\cdot)$ means $\int_S \mu(ds) R(s, \cdot)$ Also, observe that,

$$\mathbf{E}[\mathcal{N}_n(\cdot)] = \sum_{k=0}^{n-1} \mathbf{P}(Z_k \in \cdot). \tag{1.7}$$

Thus, $\frac{1}{n}\mathbf{E}[\mathcal{N}_n(\cdot)]$ is the Cesàro mean of the marginal distributions of the observed sequence of colors till time $n-1$.

1.2. *Main Achievements of the Work.* The main contributions of this work are two fold. One, we generalized urn schemes for colors indexed by an arbitrary set S and configuration of an urn is viewed as a (possibly random) measure on it. Secondly we, analyze, any such urn through two "*representations*" of the observed sequence of colors to an associated *branching Markov chain* on S with the backbone as the *random recursive tree*. These representations are novel and useful in deriving asymptotic results for the expected and random configurations of the urn.

There are few standard methods for analyzing finite color urn models which are mainly based on martingale techniques [27, 10, 11, 18], stochastic approximations [33] and embedding into continuous time pure birth processes [1, 28, 29, 3]. Typically the analysis of a finite color urn is heavily dependent on the *Perron-Frobenius theory* [42] of matrices with positive entries and *Jordan Decomposition* [17] of finite dimensional matrices [1, 27, 28, 29, 3, 10, 18]. The absence of such a theory when S is infinite, makes the analysis of urn with infinitely many colors quite difficult and challenging. In [8] the results were derived using *exchangeability* of the observed sequence of colors. However, as observed in [6], exchangeability fails in the presence of off-diagonal entries and in [6], the authors took a different approach of embedding the observed sequence of colors to the underlying random walk sequence. The major contribution of this work is to further this embedding for any general urn scheme with colors indexed by a Polish space, and then derive asymptotic results by bypassing the standard martingale and matrix theoretic techniques. As a byproduct, we also derive some non-trivial asymptotic for the *random recursive tree*(see Section 5) and generalized the work for random replacements (see Section 6).

1.3. *Limitations of the Work.* In general, for the classical as well as for the modern work on finite color urn models, typically the limits so derived are strong convergence (*almost sure*)-type [40, 27, 10, 11, 18, 1, 28, 29, 3, 33]. As we will see in the sequel, because of the very nature of our arguments, our limits will be slightly weaker (*in probability convergence*).

We conjecture here that the strong convergence should hold in general under minimal conditions on the underlying Markov chain, which should include the finite color case. One specific example has been discussed in [35] (see Section 1.4.3). While this work was under review two more work have appeared [4, 30] where strong convergence has been proved specifically for two of our examples described in the Sections 4.1.1 and 4.2.1 respectively. Both these work use the representation proved in this work and specifically one of our main result, the Theorem 2.4.

The other limitation of the work is that it only focuses on the *balanced* cases, that is, the replacement kernel R has the property that $R(s, S)$ is a constant. In other words, at each step a constant number of balls are added. Our main results, namely, Theorem 2.4 and Theorem 2.8 hold only in this case.

We certainly feel that if we consider a general *unbalanced* replacement scheme R, but assume that the function $b(s) := R(s, S)$, $s \in S$ is somewhat "*nice*" then the asymptotic results derived in this work should go through after changing the normalization from $(n+1)$ to $U_n(S)$, which will then be random. For example, we conjecture that if b is such that, $\delta \leq b(s) \leq K$ for all $s \in S$, where $\delta > 0$ and $K < \infty$. In other words, if the function b remains bounded away from 0 and ∞), then the asymptotic results derived in this work should hold after changing the normalization from $(n+1)$ to $U_n(S) = O_P(n)$.

1.4. *Discussion.* While preparing the manuscript, and after our first version was uploaded on the *arXiv* (see `https://arxiv.org/pdf/1606.05317.pdf`), we were informed by Cécile Mailler and Jean-François Marckert that they are also working on similar problems. Later their work appeared in an *arXiv* version (see `https://arxiv.org/pdf/1610.09057.pdf`) [35]. In their work, they prove more or less similar results as that of ours based on exactly the same kind of embedding which was already available in our first *arXiv* version. We here provide a brief summary of the similarities and differences of the two works.

1.4.1. *Similarities with [35].*

- Our main representation argument, stated as the *Grand Representation Theorem* (see Theorem 2.4) is exactly same to what is described in Section 2 of [35] as the coupling between their *Measure-Valued Pólya Process (MVPP)* and *Branching Markov Chain (BMC)*. However, the proof we provide is simpler and is more direct than what is given in [35].
- The main weak asymptotic result of our paper described in Theorem 3.8 are similar to that of the main asymptotic result, namely, Theorem 4 of [35]. Some of the assumptions are also equivalent, such as, our assumption (A) (see Section 3.2) is exactly same as what has been termed as $(b(n), a(n))$-ergodic in [35] (see their Definition 2).

1.4.2. *Major contributions of our work which are not in [35].*

- In Section 3, using our representation theorem we not only derive fairly general weak asymptotic for the random configuration of urn (Theorem 3.8) but we also derive weak asymptotic for the *color count statistics* under the same general conditions (Theorem 3.10). For various statistical applications this may be of importance.
- In Section 5, we provide a non-trivial application of our representation theorem to derive asymptotic of the sizes of the sub-trees rooted at the children of the root of a *random recursive tree*. This is essentially an example of a reverse application of the representation theorem.

- In Section 6, we also consider random replacement and establish corresponding representation theorems. As application we show that various non-standard limits can appear in that context for urn schemes with random replacement. These results are of somewhat different flavor than the usual random replacement matrix models studied in say [3]. However, our results derived in Section 6 partially answer the open problem stated in Section 1.6.2 of [35].
- Other than the above three, we also would like to note that our formulation works on any general color set S endowed with a σ-algebra $\mathcal{S}$. In fact, our main representation theorems, namely, Theorem 2.4 and 2.8 works in this generality. Also as pointed out in Remark 3.3 some of the convergence results also hold in full generality. In contrast, [35] needs the assumption that S is a Polish space.

1.4.3. *One major contribution of [35] which is absent in our work.*

- As mentioned above in Section 1.3, in one particular example, namely, urn associate with random walk with i.i.d. light-tail increments, the authors derives a stronger result (see Theorem 6 of [35]) of almost sure convergence for the configuration. Our Theorem 4.3 is weaker and was earlier proved in [5].

1.5. *Notations.* Most of the notations used in this paper are consistent with the literature on finite color urn models. However, we use few specific notations as well, which are given below. These are similar to what we have used earlier in [6].

(i) All vectors are written as row vectors unless otherwise stated. Column vectors are denoted by x^T, where x is a row vector.
(ii) The standard Gaussian measure on $\mathbb{R}^d$ will be denoted by Φ_d with its density given by

$$\phi_d(x) := \frac{1}{(2\pi)^{d/2}} \exp\left(-\frac{\|x\|^2}{2}\right), x \in \mathbb{R}^d.$$

For $d = 1$, we will simply write Φ for the standard Gaussian measure on $\mathbb{R}$ and ϕ for its density.
(iii) The Gaussian distribution with mean μ and variance-co-variance matrix Σ in d-dimension will be denoted by $\text{Normal}_d(\mu, \Sigma)$. For $d = 1$, we will simply write $\text{Normal}\left(\mu, \sigma^2\right)$, for $\mu \in \mathbb{R}$ and $\sigma^2 > 0$.
(iv) The symbol $\stackrel{d}{=}$ will denote equality in distribution between two random variables/stochastic processes.
(v) The symbol $\Rightarrow$ will denote convergence in distribution of random variables.
(vi) The symbol $\stackrel{p}{\longrightarrow}$ will denote convergence in probability.
(vii) The symbol $\stackrel{w}{\longrightarrow}$ will denote the weak convergence of probability measures in $\mathcal{P}(S)$.

1.6. *Outline of the Paper.* The rest of the paper is organized as follows: Section 2 contains the two representation theorems, namely, Theorem 2.4 and Theorem 2.8, which are the most important contributions of this work. In Section 3, we derive asymptotic results for random and expected configurations and the color count statistics under fairly general conditions. In Section 4, we provide many interesting applications mainly in the context of infinite color urn schemes. Section 5 provides a non-trivial application of the representation theorem for deriving certain asymptotic for the *random recursive tree*. In Section 6 we discuss random replacement and establish representation there in and few non-trivial limiting distributions of urns with random replacement. Finally, Section 7 contains few concluding remarks.

2. Main Results and Their Proofs. In this section, we present the two main theorems of this paper. These theorems, which we call the *Grand Representation Theorem* (Theorem 2.4) and the *Marginal Representation Theorem* (Theorem 2.8) provide certain "*couplings*" of the urn model with the associated Markov chain through the *observed sequence of colors*. These results are fairly general and hold for any balanced urn schemes with colors indexed by an *arbitrary* set S. Before we state the results, we introduce here two important structures, namely, *Random Recursive Tree (RRT)* and *branching Markov chain* on a RRT.

2.1. *Random Recursive Tree (RRT).* For $n \geq -1$, let $\mathcal{T}_n$ be the *random recursive tree* on $(n+2)$ vertices labeled by $\{-1; 0, 1, 2, \ldots, n\}$, where the vertex labeled by -1 is considered as the *root*. For sake of completeness we provide here the definition of the random recursive tree. A random recursive tree with $(n+2)$ vertices labeled by $\{-1; 0, 1, 2, \ldots, n\}$, is a random rooted tree, rooted at -1, and obtained by starting with a single node, labeled -1, which acts as the root and then adding $(n+1)$ vertices one by one, each time joining the new vertex to a randomly chosen existing vertex; the random choices are uniform and independent of each other. It was first introduced by Moon [37]. The survey [44] provides more details. We consider $\{\mathcal{T}_n\}_{n\geq -1}$ as a growing sequence of random trees and define

$$\mathcal{T} := \bigcup_{n\geq -1} \mathcal{T}_n, \tag{2.1}$$

and call it the *(infinite) random recursive tree*. Formally, $\mathcal{T}$ can be constructed by using a sequence, say, $(D_n)_{n\geq 0}$, of independent *discrete uniform* random variables, such that, the n-th one, namely D_n, is uniform on $\{-1; 0, 1, \cdots, n-1\}$. The n-th vertex (labeled as n), joins to a vertex labeled by the (random) index D_n. In this construction of $\mathcal{T}$, we have $\overleftarrow{n} = D_n$, where $\overleftarrow{n}$ is the parent of the n-th vertex (labeled as n). Note that in this construction the random tree $\mathcal{T}_n$ is constructed only using the variables $\{D_k\}_{k=0}^{n}$ and as n increases the random trees are growing. Such construction of random recursive trees is available in [20], and is referred to as the uniform random recursive tree in [20]. With a slight abuse of terminology, we refer to it as the random recursive tree throughout.

For $n \geq 0$, let $(1+\tau_n)$ be length of the unique path from the vertex n to the root -1 in the random recursive tree $\mathcal{T}_n$ with $n+2$ vertices, as defined above. Note $\tau_0 = 0$. The following result is known in literature (e.g. see [19, 20]). However, for the sake of completeness, we provide the proof.

LEMMA 2.1. *In the above set up, as $n \to \infty$,*

$$\frac{\tau_n}{\log n} \longrightarrow 1 \text{ a.s.} \tag{2.2}$$

and

$$\frac{\tau_n - \log n}{\sqrt{\log n}} \xrightarrow{d} N(0,1). \tag{2.3}$$

REMARK 2.2. For any vertex $v \in \mathcal{T}$, denote by $|v|_{\mathcal{T}}$ its depth from the root in $\mathcal{T}$. Observe that, the above lemma implies that if u_n is a vertex chosen uniformly at random from the vertices of $\mathcal{T}_{n-1}$, then $\frac{|u_n|_{\mathcal{T}}}{\log n} = \frac{|D_n|_{\mathcal{T}}}{\log n} \longrightarrow 1$ a.s. as $n \to \infty$. This observation we will need in the proof of Theorem 3.4.

PROOF. Now, for $0 \leq j \leq n-1$, let I_j be the indicator that the vertex j lies on the path from the root -1 to the vertex n. Then by construction $(I_j)_{0\leq j\leq n-1}$ are independent

Bernoulli variables with $\mathbf{E}\left[I_j\right] = \frac{1}{j+1}$, $0 \le j \le n-1$. Also,

$$\tau_n = \sum_{j=0}^{n-1} I_j. \tag{2.4}$$

Notice that $\mathbf{Var}\left(I_j\right) = \frac{1}{j+1}\left(1 - \frac{1}{j+1}\right)$, $0 \le j \le n-1$, thus

$$\mathbf{E}\left[\tau_n\right] = \sum_{j=0}^{n-1} \frac{1}{j+1} \sim \log n \ \text{ and } \ \mathbf{Var}\left(\tau_n\right) = \sum_{j=0}^{n-1} \frac{1}{j+1}\left(1 - \frac{1}{j+1}\right) \sim \log n, \tag{2.5}$$

as $n \to \infty$. So by Kronecker's Lemma (see (8.5) on page 63 of [21]) it follows that

$$\frac{\tau_n}{\log n} \longrightarrow 1 \text{ a.s.} \tag{2.6}$$

as $n \to \infty$, proving (2.2). Further, (2.3) follows by an easy application of the Lyapunov Central Limit Theorem (see Theorem 27.3 on page 362 of [7]). $\square$

2.2. *Branching Markov Chain on the RRT.* Let $\Delta \notin S$ be a symbol, which is outside the color set S. We define a stochastic process $\left(W_n\right)_{n \ge -1}$ taking values in $\hat{S} := \{\Delta\} \cup S$, such that, $W_{-1} = \Delta$, and for any $n \ge 0$ and $A \in \mathcal{S}$,

$$\mathbf{P}\left(W_n \in A \,\middle|\, W_{n-1}, W_{n-2}, \ldots, W_{-1}; \mathcal{T}_n\right) = \begin{cases} U_0\left(A\right) & \text{if } W_{\overleftarrow{n}} = \Delta; \\ R\left(W_{\overleftarrow{n}}, A\right) & \text{otherwise}, \end{cases} \tag{2.7}$$

where recall $\overleftarrow{n}$ is the parent of the vertex labeled by n. This process, namely, $\left\{W_n\right\}_{n \ge 0}$ will be called a *branching Markov chain* on the RRT, starting at the root -1 and at a position $W_{-1} = \Delta$ with the Markov kernel $\hat{R}$ on $\hat{S}$, defined as,

$$\hat{R}\left(w, \cdot\right) = \begin{cases} U_0\left(\cdot\right) & \text{if } w = \Delta; \\ R\left(w, \cdot\right) & \text{otherwise}. \end{cases} \tag{2.8}$$

The following results are immediate consequence of the definition.

PROPOSITION 2.3. Let $\left\{W_n\right\}_{n \ge 0}$ be a *branching Markov chain* on the RRT $\mathcal{T}$, as defined above. Then for any $n \ge 0$, if $v_0 = -1; v_1, \cdots, v_{\tau_n} = \overleftarrow{n}, v_{1+\tau_n} = n$ be the *unique* path from the root -1 to the vertex n in the RRT $\mathcal{T}_n$ on $(n+2)$-vertices, then given $\mathcal{T}_n$, and on the event $\left[\tau_n = t\right]$, the conditional law of the chain $W_{v_1}, W_{v_2}, \cdots, W_{v_{1+t}}$ is that of a Markov chain on length t on S starting with the initial distribution U_0 and Markov kernel R.

2.3. *Grand Representation Theorem.* The following theorem is a "*representation*" of the entire sequence of colors $\left(Z_n\right)_{n \ge 0}$ in terms of the *Branching Markov chain on random recursive tree*.

THEOREM 2.4. *Consider an urn model with colors indexed by a set S endowed with a σ-algebra $\mathcal{S}$. Let R be the replacement kernel and U_0 be the initial configuration. For $n \ge 0$, let Z_n be the random color of the $(n+1)$-th draw. Let $\left(W_n\right)_{n \ge -1}$ be the* branching Markov chain *on $\mathcal{T}$ as defined in Section 2.2. Then*

$$\left(Z_n\right)_{n \ge 0} \stackrel{d}{=} \left(W_n\right)_{n \ge 0}. \tag{2.9}$$

PROOF. We will prove the result by establishing a *coupling* between the branching Markov chain $(W_n)_{n\geq 0}$ defined on the (infinite) random recursive tree $\mathcal{T}$, and the random drawn color sequence, namely, $(Z_n)_{n\geq 0}$.

We start by observing that from (1.1), we get for $n \geq 1$,

$$\frac{U_n}{n+1} = \frac{U_0}{n+1} + \sum_{k=0}^{n-1} \frac{R(Z_k, \cdot)}{n+1}. \tag{2.10}$$

Also recall, that given the random configuration of the urn U_n, the $(n+1)$-th drawn color, namely, Z_n has the (conditional) distribution given by $\frac{U_n}{n+1}$. Thus, using the equation (2.10) we conclude that

$$\mathbf{P}\left(Z_n \in \cdot \,\middle|\, Z_{n-1}, Z_{n-2}, \cdots, Z_1, Z_0, U_0\right) = \frac{U_0(\cdot)}{n+1} + \sum_{k=0}^{n-1} \frac{R(Z_k, \cdot)}{n+1}. \tag{2.11}$$

Now let $(D_n)_{n\geq 0}$ be the sequence of random variables as defined in Section 2.1 and we construct the (infinite) random recursive tree $\mathcal{T}$ using them and also the branching random walk $(W_n)_{n\geq -1}$ on it as described above. Then for any $n \geq 0$ and $A \in \mathcal{S}$,

$$\mathbf{P}\left(W_n \in A \,\middle|\, W_{n-1}, W_{n-2}, \cdots, W_1, W_0; U_0, \mathcal{T}_n\right) = \hat{R}\left(W_{\overleftarrow{n}}, A\right). \tag{2.12}$$

But since $\overleftarrow{n} \overset{d}{=} D_n \sim \text{Uniform}\{-1; 0, 1, \cdots, n-1\}$, so we get

$$\begin{aligned}
&\mathbf{P}\left(W_n \in A \,\middle|\, W_{n-1}, W_{n-2}, \cdots, W_1, W_0; U_0\right) \\
&= \mathbf{E}\left[\hat{R}\left(W_{\overleftarrow{n}}, A\right) \,\middle|\, W_{n-1}, W_{n-2}, \cdots, W_1, W_0; U_0\right] \\
&= \mathbf{E}\left[\mathbf{1}_{\left[\overleftarrow{n}=-1\right]} U_0(A) + \mathbf{1}_{\left[\overleftarrow{n}\geq 0\right]} R\left(W_{\overleftarrow{n}}, A\right) \,\middle|\, W_{n-1}, W_{n-2}, \cdots, W_1, W_0; U_0\right] \\
&= \frac{U_0(A)}{n+1} + \sum_{k=0}^{n-1} \frac{R(W_k, A)}{n+1}. \qquad (2.13)
\end{aligned}$$

Finally, observing that $Z_0 \sim U_0$ and $W_0 \sim U_0$, and using equations (2.11) and (2.13) we conclude that (2.9) follows by induction. $\square$

Following corollary follows immediately from the proof above.

COROLLARY 2.5. Consider an urn model with colors indexed by a set S endowed with a σ-algebra $\mathcal{S}$. Let R be the replacement kernel and U_0 be the initial configuration. For $n \geq 0$, let Z_n be the random color of the $(n+1)$-th draw. Let $(W_n)_{n\geq -1}$ be the *branching Markov chain* on $\mathcal{T}$ as defined in Section 2.2. Then

$$\left(\frac{U_n(\cdot)}{n+1}\right)_{n\geq 0} \overset{d}{=} \left(\mathbf{P}\left(W_n \in \cdot \,\middle|\, W_{n-1}, W_{n-2}, \cdots, W_1, W_0; U_0\right)\right)_{n\geq 0}. \tag{2.14}$$

REMARK 2.6. The last result essentially states that the sequence of the random configurations of the urn, namely, $\left(\frac{U_n}{n+1}\right)_{n\geq 0}$, can in principle be studied by observing only the branching Markov chain variables, namely, $(W_n)_{n\geq -1}$. This is a powerful relation which we will make use in the rest of the paper.

Our next result states the relation between the *color count statistics* and the branching Markov chain.

COROLLARY 2.7. Consider an urn model with colors indexed by a set S endowed with a σ-algebra $\mathcal{S}$. Let R be the replacement kernel and U_0 be the initial configuration. For $n \geq 0$, let Z_n be the random color of the $(n+1)$-th draw. Let $(W_n)_{n\geq -1}$ be the *branching Markov chain* on $\mathcal{T}$ as defined in Section 2.2. Then

$$(\mathcal{N}_n)_{n\geq 1} \stackrel{d}{=} \left(\sum_{k=0}^{n-1} \delta_{W_k}\right)_{n\geq 1}. \tag{2.15}$$

2.4. *Marginal Representation Theorem.* Our next result is a "*representation*" of the marginal distribution for the randomly chosen color Z_n in terms of the marginal distribution of the corresponding Markov chain sampled at random but independent times. As we will see from the proof it is an immediate consequence of Theorem 2.4.

THEOREM 2.8. *Consider an urn model with colors indexed by a set S endowed with a σ-algebra $\mathcal{S}$. Let R be the replacement kernel and U_0 be the initial configuration. For $n \geq 0$, let Z_n be the random color of the $(n+1)$-th draw. Then there exist a Markov chain $(X_n)_{n\geq 0}$ on S with transition kernel R and initial distribution U_0 and an increasing sequence of random indices $(\tau_n)_{n\geq 0}$ with $\tau_0 = 0$, which are independent of the Markov chain $(X_n)_{n\geq 0}$, such that,*

$$Z_n \stackrel{d}{=} X_{\tau_n}, \tag{2.16}$$

for any $n \geq 0$. Moreover, the sequence of random indices $(\tau_n)_{n\geq 0}$ satisfies equations (2.2) and (2.3).

REMARK 2.9. A version of this last result was obtained in Proposition 7 in [6], which was restricted to the case when $(X_n)_{n\geq 0}$ is a bounded increment random walk. Here however, the result is for any general Markov chain $(X_n)_{n\geq 0}$.

REMARK 2.10. It is worthwhile to note here that, it is not necessary that the law of the sequence $(Z_n)_{n\geq 0}$ is same as the law of $(X_{\tau_n})_{n\geq 0}$, where the random variables are as defined in Theorem 2.8. This is because $(Z_n)_{n\geq 0}$ is not necessarily Markov, but $(X_{\tau_n})_{n\geq 0}$ is necessarily is a Markovian sequence. In fact, the law of the process $(Z_n)_{n\geq 0}$ is more complicated as presented in Theorem 2.4.

PROOF. As before, let $(1 + \tau_n)$ be length of the unique path from the vertex n to the root -1 in the random recursive tree $\mathcal{T}_n$ with $n + 2$ vertices. Thus the sequence of random variables $(\tau_n)_{n\geq 0}$ satisfy the equations (2.2) and (2.3).

Now, on the same probability space where $(W_n)_{n\geq -1}$ and $(\mathcal{T}_n)_{n\geq -1}$ are defined, construct a Markov chain $(X_n)_{n\geq 0}$ on S, starting at $X_0 \sim U_0$ with Markov kernel R which is independent of $(W_n)_{n\geq -1}$ and $(\mathcal{T}_n)_{n\geq -1}$. So the sequence of $(\tau_n)_{n\geq 0}$ is independent of the Markov chain $(X_n)_{n\geq 0}$ by construction.

Using Proposition 2.3, for any $n \geq 0$, $A \in \mathcal{S}$ and $t \geq 0$,

$$\mathbf{P}\left(W_n \in A \,\middle|\, \mathcal{T}_n, \tau_n = t\right) = \mathbf{P}\left(X_t \in A\right) \mathbf{1}_{\tau_n = t} \tag{2.17}$$

Thus by taking expectation and summing over t, we get

$$W_n \stackrel{d}{=} X_{\tau_n}, \tag{2.18}$$

and thus (2.16) follows from Theorem 2.4. $\square$

COROLLARY 2.11. Consider an urn model with colors indexed by a set S endowed with a σ-algebra $\mathcal{S}$. Let R be the replacement kernel and U_0 be the initial configuration. For $n \geq 0$, let Z_n be the random color of the $(n+1)$-th draw. Then there exist a Markov chain $(X_n)_{n\geq 0}$ on S with transition kernel R and initial distribution U_0 and an increasing sequence of random indices $(\tau_n)_{n\geq 0}$ with $\tau_0 = 0$, which are independent of the Markov chain $(X_n)_{n\geq 0}$, such that,

$$\frac{\mathbf{E}\left[U_n\left(\cdot\right)\right]}{n+1} = \mathbf{P}\left(X_{\tau_n} \in \cdot\right). \tag{2.19}$$

PROOF. Recall that from (1.4) the probability mass function of Z_n is given by $\frac{1}{n+1}\mathbf{E}\left[U_n\right]$. Thus, the equation (2.19) holds by using the Theorem 2.8. □

3. Weak Asymptotic of the Urn Configuration and the Color Count Statistics. In this section we state and prove some very general results for the asymptotic of the random and expected configurations and the color count statistics of our general urn scheme $(U_n)_{n\geq 0}$, defined in Section 1.1. These results will be proved using the two representations theorems given in Section 2. We start by establishing an asymptotic result for the branching Markov chain $(W_n)_{n\geq -1}$ as defined in the Section 2.3. For this and the later sections, we assume that $\mathcal{P}\left(S\right)$ is endowed with the topology of *weak convergence* and any limit statement in $\mathcal{P}\left(S\right)$ is with respect to the topology of weak convergence.

Let us first recall that $(W_n)_{n\geq -1}$ is defined as the *branching Markov chain* on the (infinite) random recursive tree $\mathcal{T} := \cup_{n\geq -1}\mathcal{T}_n$, starting at the root -1 and at a position $\Delta \notin S$. Define $\mathcal{G}_n := \sigma\left(W_0, W_1, \cdots, W_{n-1}\right)$, $n \geq 0$. Let Q_n be a version of the regular conditional distribution of W_n given $\mathcal{G}_n$. Note that Q_n exists and is almost surely unique and proper, as S is a Polish space and $\mathcal{S}$ is the corresponding Borel σ-algebra. In fact, almost surely,

$$Q_n\left(\cdot\right) = \frac{1}{n+1}\sum_{m=-1}^{n-1}\hat{R}\left(W_m, \cdot\right), \tag{3.1}$$

where $\hat{R}$ is defined in (2.8) and $W_{-1} = \Delta$, as defined in Section 2.2. Further, if q_n be a version of the regular conditional distribution of W_n given $\mathcal{G}_n$ and $\mathcal{T}_{n-1}$, then because of the similar reason as above q_n exists and is almost surely unique and proper. But, it is immediate that almost surely,

$$q_n\left(\cdot\right) = \frac{1}{n+1}\sum_{m=-1}^{n-1}\hat{R}\left(W_m, \cdot\right). \tag{3.2}$$

Thus,

$$Q_n = q_n \text{ a.s.} \tag{3.3}$$

This observation also tells us that condition on $\mathcal{G}_n$ the variable W_n is independent of the random tree $\mathcal{T}_{n-1}$.

Also let, $\mathcal{E}_n$ be the (random) *empirical measure* of the variables $(W_k)_{k=0}^{n-1}$, that is,

$$\mathcal{E}_n := \frac{1}{n}\sum_{k=0}^{n-1}\delta_{W_k}. \tag{3.4}$$

Notice that

$$Q_n = \frac{U_0}{n+1} + \frac{n}{n+1}\mathcal{E}_n R. \tag{3.5}$$

Note that similar relationship holds for the sequence of random measures $\left(\frac{U_n}{n+1}\right)_{n\geq 0}$ and $(\mathcal{N}_n)_{n\geq 1}$, see equation (1.6)), which is consistent with Corollary 2.7.

It is worth noting here that by Corollary 2.5, $(Q_n)_{n\geq 0}$ corresponds to $\left(\frac{U_n}{n+1}\right)_{n\geq 0}$. And, by Corollary 2.7 $(\mathcal{E}_n)_{n\geq 1}$ corresponds to $\left(\frac{\mathcal{N}_n}{n}\right)_{n\geq 1}$.

3.1. *An Assumption.* Recall that $(X_n)_{n\geq 0}$ denotes a Markov chain with state space S, transition kernel R and starting distribution U_0 and $S \subseteq \mathbb{R}^d$ for some $d \geq 1$. We now make the following assumption:

(A) There exists a (non-random) probability Λ on $\left(\mathbb{R}^d, \mathcal{B}_{\mathbb{R}^d}\right)$ and a vector $\mathbf{v} \in \mathbb{R}^d$, and two functions $a : \mathbb{R}_+ \to \mathbb{R}$ and $b : \mathbb{R}_+ \to \mathbb{R}_+$, such that, for any initial distribution U_0,

$$\frac{X_n - a(n)\,\mathbf{v}}{b(n)} \Rightarrow \Lambda \text{ in } \mathcal{P}\left(\mathbb{R}^d\right). \tag{3.6}$$

REMARK 3.1. At first glance the assumption made above may look fairly restrictive. However, it is satisfied by a large class of interesting examples, including the well studied classical urn model case [27, 28], where R represents a finite state irreducible, aperiodic Markov chain. More examples are discussed in Section 4. In fact, in some sense the assumption above essentially says that the associated Markov chain is "*ergodic*". This is because we assume that after appropriate centering and scaling the chain has a limiting distribution which is independent of the starting distribution. It is in fact, hard to get examples of Markov chains taking values in a finite dimensional Euclidean space, which is "*irreducible*" (see [36] for a formal definition), but the assumption (A) does not hold. Examples may be obtained from *Random Walks in Random Environment (RWRE)*. We discuss such cases in Section 6

REMARK 3.2. We emphasized here that Assumption (A) implies that for any $s \in S$,

$$\mathbf{P}\left(\frac{X_n - a(n)\,\mathbf{v}}{b(n)} \in \cdot \,\middle|\, X_0 = s\right) \longrightarrow \Lambda(\cdot). \tag{3.7}$$

This is because Assumption (A) holds for every initial distribution and also the limiting distribution is non-random. This fact is used in the proofs later.

REMARK 3.3. The assumption that S is a Polish space and is a subset of $\mathbb{R}^d$ is only necessary for this assumption (A) to go through when a and b are non-trivial. If $a = 0$ and $b = 1$ then assumption (A) can hold for any general state space S endowed with a sigma algebra $\mathcal{S}$ and in that case the assumption (A) should be read as:

There exists a (non-random) probability Λ on $(S, \mathcal{S})$ such that, for any initial distribution U_0,

$$X_n \Rightarrow \Lambda \text{ in } \mathcal{P}(S).$$

Thus Part (a) of the Theorems 3.4, 3.7, 3.8, 3.9, and 3.10 hold without any Polish space assumption on S and/or any embedding into an Euclidean space. The examples discussed in Section 4.1 also works in this generality.

3.2. *Asymptotic of Branching Markov Chain on Random Recursive Tree.* We now prove certain weak asymptotic for the branching Markov chain when the associate Markov chain satisfy the assumption (A).

THEOREM 3.4. *Suppose that the assumption **(A)** holds. Let $Q_n^{cs(r)}$ be the conditional distribution of $\dfrac{W_n - a(1+\tau_n)\mathbf{v}}{b(1+\tau_n)}$ given $\mathcal{G}_n$, that is, a centered and scaled version of Q_n with (possibly random) centering by $a(1+\tau_n)\mathbf{v}$ and scaling by $b(1+\tau_n)$. Then under the assumptions stated in (a), (b) & (c) below and as $n \to \infty$,*

$$Q_n^{cs(r)} \xrightarrow{p} \Lambda, \tag{3.8}$$

where the above convergence is in $\mathcal{P}(S)$ under the conditions given in (a) below, otherwise in $\mathcal{P}\left(\mathbb{R}^d\right)$. Moreover, let Q_n^{cs} be the conditional distribution of $\dfrac{W_n - a(\log n)\mathbf{v}}{b(\log n)}$ given $\mathcal{G}_n$, that is, a centered and scaled version of Q_n with (non-random) centering by $a(\log n)\mathbf{v}$ and scaling by $b(\log n)$, then

(a) *If $a = 0$ and $b = 1$, then*

$$Q_n^{cs} = Q_n \xrightarrow{p} \Lambda \text{ in } \mathcal{P}(S). \tag{3.9}$$

(b) *Suppose $a = 0$ and b is regularly varying function, then*

$$Q_n^{cs} \xrightarrow{p} \Lambda \text{ in } \mathcal{P}\left(\mathbb{R}^d\right). \tag{3.10}$$

(c) *Suppose a is differentiable and $\lim_{x\to\infty} a'(x) = \tilde{a} < \infty$. Also assume b is regularly varying and $\lim_{x\to\infty} \dfrac{\sqrt{x}}{b(x)} = \tilde{b} < \infty$ then*

$$Q_n^{cs} \xrightarrow{p} \Xi \text{ in } \mathcal{P}\left(\mathbb{R}^d\right), \tag{3.11}$$

where Ξ is Λ if $\tilde{a} = 0$ or $\tilde{b} = 0$, otherwise, it is given by the convolution of Λ and Normal $\left(0, \tilde{a}^2\tilde{b}^2\right)\mathbf{v}$.

REMARK 3.5. It is worthwhile to note here that the asymptotic limit of Q_n^{cs} for Parts (a) and (b) follow almost immediately from the limiting distribution of $Q_n^{cs(r)}$ and the limit remains same, namely, Λ, which is the scaled limit in this cases for the associated Markov chain. For Part (c) above where there is a non-trivial centering for Q_n^{cs} a possibly different limit is obtained, which is a random but independent Gaussian shift of Λ. As seen in the proof given below, this random Gaussian shift appears due to the non-trivial centering term which depends on the random recursive tree and the centering and scaling functions a and b.

PROOF. We start by proving that under the assumption **(A)**, the equation (3.8) holds. For this we break the proof in several steps as given below.

Proof of Equation (3.8). STEP I: To show that $\mathbf{E}\left[Q_n^{cs(r)}(\cdot)\right] \longrightarrow \Lambda(\cdot)$ in $\mathcal{P}\left(\mathbb{R}^d\right)$:

We begin by defining a new measure, which we denote by $\tilde{q}_n^{cs}$. It is the conditional distribution of $\dfrac{W_n - a(1+\tau_n)\mathbf{v}}{b(1+\tau_n)}$ given $\mathcal{T}_n$. From definition it follows:

$$\mathbf{E}\left[Q_n^{cs(r)}(\cdot)\right] = \mathbf{E}\left[\tilde{q}_n^{cs}(\cdot)\right]. \tag{3.12}$$

Thus it is enough to show $\mathbf{E}\left[\tilde{q}_n^{cs}(\cdot)\right] \longrightarrow \Lambda(\cdot)$ in $\mathcal{P}\left(\mathbb{R}^d\right)$.

Let ρ be a metric on $\mathcal{P}(S)$, which metrizes the weak convergence topology on it. Denote by L_n the distribution of $\frac{X_n - a(n)}{b(n)}$. Under assumption **(A)**, we have

$$\rho\left(L_n, \Lambda\right) \longrightarrow 0, \tag{3.13}$$

as $n \to \infty$.

Now, fix $\epsilon > 0$ and find $H > 0$ so large that $\rho\left(L_h, \Lambda\right) < \epsilon$, for any $h > H$. Find $N \geq 1$, so large that $\frac{(\log(n+2))^H}{n+2} < \epsilon$ for all $n \geq N$.

Recall that, D_n denotes the vertex at which n-th vertex joins in the random recursive tree $\mathcal{T}_n$. So from definition,

$$\tilde{q}_n\left(d\mathbf{x}\right) = \mathbf{E}\left[\hat{R}\left(W_{D_n}, a\left(1+\tau_n\right) + b\left(1+\tau_n\right) d\mathbf{x}\right) \,\Big|\, \mathcal{T}_n\right].$$

Thus, given $\mathcal{T}_n$,

$$\rho\left(\tilde{q}_n^{cs}, \Lambda\right) \stackrel{d}{=} \sum_{j=0}^{n} \mathbf{1}_{[\tau_n = j]}\, \rho\left(L_{j+1}, \Lambda\right), \tag{3.14}$$

where we recall, that $1 + \tau_n$ is the length of the unique path from the vertex n to the root $\mathcal{T}_n$.

Now let S_n^H be the set of vertices of the random recursive tree $\mathcal{T}_n$ up to depth H. Then,

$$\begin{aligned}
&\mathbf{P}\left(\rho\left(\tilde{q}_n^{cs}, \Lambda\right) > \epsilon\right) \\
&= \mathbf{E}\left[\mathbf{P}\left(\sum_{j=0}^{n} \mathbf{1}_{[\tau_n = j]}\, \rho\left(L_{j+1}, \Lambda\right) > \epsilon \,\Big|\, \mathcal{T}_n\right)\right] \\
&\leq \mathbf{P}\left(D_n \in S_n^H\right) \\
&= \frac{\mathbf{E}\left[\left|S_n^H\right|\right]}{n+2} < \epsilon,
\end{aligned} \tag{3.15}$$

where H is as chosen above and the last inequality follows from the Lemma 3.6. This completes the proof of the STEP I.

In fact, it is worth noting here that what we proved above is indeed,

$$\tilde{q}_n^{cs} \stackrel{p}{\longrightarrow} \Lambda\left(\cdot\right) \ \text{ in } \mathcal{P}\left(\mathbb{R}^d\right). \tag{3.16}$$

Now let $C\left(\Lambda\right) \subseteq \mathbb{R}^d$ be the set of all points of continuity of the measure Λ. Also for two vectors $\mathbf{x}, \mathbf{y} \in \mathbb{R}^d$ we will write $\mathbf{x} \leq \mathbf{y}$, if and only if, the inequalities hold component wise. Denote by $\left(-\infty, \mathbf{x}\right] = \prod_{i=1}^{d}\left(-\infty, x_i\right]$, where $\mathbf{x} := \left(x_1, x_2, \cdots, x_d\right)$.

STEP II: To show that $\mathbf{Var}\left(Q_n^{cs(r)}\left(-\infty, \mathbf{x}\right]\right) \longrightarrow 0 \ \forall \mathbf{x} \in C\left(\Lambda\right)$:

Fix $\mathbf{x} \in C\left(\Lambda\right) \subseteq \mathbb{R}^d$. Recall, that from definition

$$Q_n^{cs(r)}\left(-\infty, \mathbf{x}\right] = \frac{1}{n+1} \sum_{m=-1}^{n-1} \hat{R}\left(W_m, \left(-\infty, a\left(1 + |m|_{\mathcal{T}}\right)\mathbf{v} + b\left(1 + |m|_{\mathcal{T}}\right)\mathbf{x}\right]\right) \tag{3.17}$$

where in the last equality we write $|m|_{\mathcal{T}}$ as the distance of the vertex m from the root -1 in the RRT $\mathcal{T}$. Also note that for $-1 \leq m \leq n-1$, the random variable $|m|_{\mathcal{T}}$ is measurable with respect to $\mathcal{T}_{n-1}$.

Now let $\left(u_n, v_n\right)$ be two vertices chosen uniformly at random from the set of vertices of $\mathcal{T}_{n-1}$. Construct two random variables $W_{u_n}^+$ and $W_{v_n}^+$ on the same probability space, such that, given $\mathcal{G}_n$ and (u_n, v_n), the two random variables $W^+_{u_n}$ and $W^+_{v_n}$ are independent and have distributions given by $\hat{R}(W_{u_n}, \cdot)$ and $\hat{R}(W_{v_n}, \cdot)$ respectively. Then

$$\begin{aligned}
&\mathbf{Var}\left(Q_n^{cs(r)}\left((-\infty, \mathbf{x}]\right)\right)\\
&= \frac{1}{(n+1)^2}\sum_{m,j=-1}^{n-1}\mathbf{Cov}\Big(\hat{R}\left(W_m, \left(-\infty, a\left(1+|m|_{\mathcal{T}}\right)\mathbf{v}+b\left(1+|m|_{\mathcal{T}}\right)\mathbf{x}\right]\right),\\
&\qquad\qquad \hat{R}\left(W_j, \left(-\infty, a\left(1+|j|_{\mathcal{T}}\right)\mathbf{v}+b\left(1+|j|_{\mathcal{T}}\right)\mathbf{x}\right]\right)\Big)\\
&= \mathbf{Cov}\left(\mathbf{1}\left(\frac{W^+_{u_n}-a\left(1+|u_n|_{\mathcal{T}}\right)\mathbf{v}}{b\left(1+|u_n|_{\mathcal{T}}\right)} \le \mathbf{x}\right), \mathbf{1}\left(\frac{W^+_{v_n}-a\left(1+|v_n|_{\mathcal{T}}\right)\mathbf{v}}{b\left(1+|v_n|_{\mathcal{T}}\right)} \le \mathbf{x}\right)\right)\\
&= \mathbf{P}\left(\frac{W^+_{u_n}-a\left(1+|u_n|_{\mathcal{T}}\right)\mathbf{v}}{b\left(1+|u_n|_{\mathcal{T}}\right)} \le \mathbf{x} \text{ and } \frac{W^+_{v_n}-a\left(1+|v_n|_{\mathcal{T}}\right)\mathbf{v}}{b\left(1+|v_n|_{\mathcal{T}}\right)} \le \mathbf{x}\right)\\
&\qquad -\left(\mathbf{P}\left(\frac{W^+_{u_n}-a\left(1+|u_n|_{\mathcal{T}}\right)\mathbf{v}}{b\left(1+|u_n|_{\mathcal{T}}\right)} \le \mathbf{x}\right)\right)^2
\end{aligned}\tag{3.18}$$

Now the second term in the above equation is exactly same as $\left(\mathbf{P}\left(\frac{W_n-a(1+\tau_n)\mathbf{v}}{b(1+\tau_n)} \le \mathbf{x}\right)\right)^2$. Thus, converges to $\Lambda^2\left((-\infty, \mathbf{x}]\right)$ by Step I. So it is enough to show that the first term also converges to $\Lambda^2\left((-\infty, \mathbf{x}]\right)$.

Now let ξ_n be the *least common ancestor* of the vertices u_n and v_n in $\mathcal{T}_{n-1}$. Then

$$\begin{aligned}
&\mathbf{P}\left(\frac{W^+_{u_n}-a\left(1+|u_n|_{\mathcal{T}}\right)\mathbf{v}}{b\left(1+|u_n|_{\mathcal{T}}\right)} \le \mathbf{x} \text{ and } \frac{W^+_{v_n}-a\left(1+|v_n|_{\mathcal{T}}\right)\mathbf{v}}{b\left(1+|v_n|_{\mathcal{T}}\right)} \le \mathbf{x}\right)\\
&= \mathbf{E}\left[\mathbf{P}\left(\frac{W^+_{u_n}-a\left(1+|u_n|_{\mathcal{T}}\right)\mathbf{v}}{b\left(1+|u_n|_{\mathcal{T}}\right)} \le \mathbf{x} \text{ and } \frac{W^+_{v_n}-a\left(1+|v_n|_{\mathcal{T}}\right)\mathbf{v}}{b\left(1+|v_n|_{\mathcal{T}}\right)} \le \mathbf{x} \,\middle|\, \xi_n, W_{\xi_n}, \mathcal{T}\right)\right]
\end{aligned}\tag{3.19}$$

Now, given $[|\xi_n|_{\mathcal{T}} = k, W_k, \mathcal{T}]$, the distribution of the variable $W^+_{u_n}$ is same as the distribution of a Markov chain starting at δ_{W_k} with replacement kernel R and have taken a total of $(1+|u_n|_{\mathcal{T}}-k)$-many steps. Similar arguments follow for $W^+_{v_n}$ and they are independent. So from (3.19) we get

$$\begin{aligned}
&\mathbf{P}\left(W^+_{u_n} \le a\left(1+|u_n|_{\mathcal{T}}\right)\mathbf{v}+b\left(1+|u_n|_{\mathcal{T}}\right)\mathbf{x} \text{ and } W^+_{v_n} \le a\left(1+|v_n|_{\mathcal{T}}\right)\mathbf{v}+b\left(1+|v_n|_{\mathcal{T}}\right)\mathbf{x}\right)\\
&= \sum_{k=0}^{\infty}\mathbf{P}\left(|\xi_n|_{\mathcal{T}}=k\right)\int_S\int_S \hat{R}^k(\Delta, dt)\mathbf{E}\Big[\hat{R}^{1+|u_n|_{\mathcal{T}}-k}\left(t, \left(-\infty, a\left(1+|u_n|_{\mathcal{T}}\right)\mathbf{v}+b\left(1+|u_n|_{\mathcal{T}}\right)\mathbf{x}\right]\right)\\
&\qquad \times\hat{R}^{1+|v_n|_{\mathcal{T}}-k}\left(t, \left(-\infty, a\left(1+|v_n|_{\mathcal{T}}\right)\mathbf{v}+b\left(1+|v_n|_{\mathcal{T}}\right)\mathbf{x}\right]\right)\Big]
\end{aligned}\tag{3.20}$$

Now, from **Theorem 7** of [32] we know that the sequence $\left(|\xi_n|_{\mathcal{T}}\right)_{n\ge 0}$ converges weakly to Geometric(1.2) -distribution. In particular, it remains *tight*. Thus given $\epsilon > 0$ there exists $N \ge 0$, such that, $\mathbf{P}\left(|\xi_n|_{\mathcal{T}} \le N\right) > 1-\epsilon$ for all $n \ge 0$

So from (3.20) we get

$$\begin{aligned}
&\Bigg|\mathbf{P}\left(\frac{W^+_{u_n}-a\left(1+|u_n|_{\mathcal{T}}\right)\mathbf{v}}{b\left(1+|u_n|_{\mathcal{T}}\right)} \le \mathbf{x} \text{ and } \frac{W^+_{v_n}-a\left(1+|v_n|_{\mathcal{T}}\right)\mathbf{v}}{b\left(1+|v_n|_{\mathcal{T}}\right)} \le \mathbf{x}\right)\\
&- \sum_{k=0}^{N}\mathbf{P}\left(|\xi_n|_{\mathcal{T}}=k\right)\int_S\int_S \hat{R}^k(\Delta, dt)\mathbf{E}\Big[\hat{R}^{1+|u_n|_{\mathcal{T}}-k}\left(t, \left(-\infty, a\left(1+|u_n|_{\mathcal{T}}\right)\mathbf{v}+b\left(1+|u_n|_{\mathcal{T}}\right)\mathbf{x}\right]\right)
\end{aligned}$$

$$\times \hat{R}^{1+|v_n|_{\mathcal{T}}-k}\left(t,\left(-\infty, a\left(1+|v_n|_{\mathcal{T}}\right)\mathbf{v}+b\left(1+|v_n|_{\mathcal{T}}\right)\mathbf{x}\right]\right)\Big]\Bigg| \le \epsilon \tag{3.21}$$

Further, from Lemma 2.1 it follows that for every fixed $k \ge 1$, $(1+|u_n|_{\mathcal{T}}-k) \longrightarrow \infty$ and $(1+|v_n|_{\mathcal{T}}-k) \longrightarrow \infty$ almost surely. Thus, under the assumptions given in the parts (a), (b) & (c), we can conclude that each of the terms appearing within the expectation sign inside the summation sign of the equation (3.21) above, converges to $\Lambda\left(\left(-\infty,\mathbf{x}\right]\right)$ almost surely. Finally, using DCT we can conclude that

$$\mathbf{P}\left(\frac{W^+_{u_n}-a\left(1+|u_n|_{\mathcal{T}}\right)\mathbf{v}}{b\left(1+|u_n|_{\mathcal{T}}\right)} \le \mathbf{x} \text{ and } \frac{W^+_{v_n}-a\left(1+|v_n|_{\mathcal{T}}\right)\mathbf{v}}{b\left(1+|v_n|_{\mathcal{T}}\right)} \le \mathbf{x}\right) \longrightarrow \Lambda^2\left(\left(-\infty,\mathbf{x}\right]\right). \tag{3.22}$$

This completes the proof of the Step II.

STEP III: To show that equation (3.8) holds:

Now using the STEPS I & II we get that

$$Q_n^{cs(r)}\left(-\infty,\mathbf{x}\right] \xrightarrow{p} \Lambda\left(-\infty,\mathbf{x}\right] \quad \forall \mathbf{x} \in C\left(\Lambda\right). \tag{3.23}$$

Thus by a Cantor-type *diagonal argument* it follows that given any sub-sequence $\{n_k\}_{k=1}^{\infty}$ there exists a further sub-sequence $\left\{n_{k_j}\right\}_{j=1}^{\infty}$, such that,

$$Q_n^{cs(r)}\left(-\infty,\mathbf{x}\right] \longrightarrow \Lambda\left(-\infty,\mathbf{x}\right] \quad \forall \mathbf{x} \in \mathbb{Q}^d \cap C\left(\Lambda\right) \text{ a.s.}$$

But as $\mathbb{Q}^d$ is dense in $\mathbb{R}^d$, it follows that given any sub-sequence $\{n_k\}_{k=1}^{\infty}$ there exists a further sub-sequence $\left\{n_{k_j}\right\}_{j=1}^{\infty}$, such that,

$$Q_n^{cs(r)}\left(-\infty,\mathbf{x}\right] \longrightarrow \Lambda\left(-\infty,\mathbf{x}\right] \quad \forall \mathbf{x} \in C\left(\Lambda\right) \text{ a.s.}$$

This proves that equation (3.8) holds.

Proof of Part (a). It is enough to observe that under the assumptions of $a = 0$ and $b = 1$, $Q_n^{cs}\left(\cdot\right) = Q_n^{cs(r)}\left(\cdot\right)$ and hence the proof follows from (3.8).

Proof of Part (b). We first observe that if b is regularly varying, then by Karamata's Characterization Theorem [26] and equation (2.6) we can show that

$$\frac{b\left(\log n\right)}{b\left(1+\tau_n\right)} \xrightarrow{p} 1. \tag{3.24}$$

Now observe that

$$\frac{W_n}{b\left(\log n\right)} = \frac{b\left(1+\tau_n\right)}{b\left(\log n\right)}\frac{W_n}{b\left(1+\tau_n\right)}. \tag{3.25}$$

Now using (3.8), and the equations (3.24) and (3.25) and applying the *Converging Together Lemma* (also known as, the *Slutsky's Theorem*) (see Exercise 2.10 in Section 2.2 of [21]) we conclude that (3.10) holds.

Proof of Part (c). We first note that under the assumptions made in (c), using the standard *Mean Value Theorem* (see Theorem 5.10 on pg. 108 of [41]) we get

$$a\left(1+\tau_n\right) - a\left(\log n\right) = a'\left(\eta_n\right)\left(1+\tau_n - \log n\right) \tag{3.26}$$

where

$$\min\left(1+\tau_n, \log n\right) \le \eta_n \le \max\left(1+\tau_n, \log n\right). \tag{3.27}$$

So from (2.2) we get that

$$\frac{\eta_n}{\log n} \longrightarrow 1 \text{ a.s.} \tag{3.28}$$

and hence

$$a'(\eta_n) \longrightarrow \tilde{a} \text{ a.s.} \tag{3.29}$$

The last conclusion is because of the assumptions made in (c). Thus using equations (2.3) and (3.29) and applying again the Converging Together Lemma, we get

$$\frac{a(1+\tau_n) - a(\log n)}{\sqrt{\log n}} \Rightarrow \begin{cases} \text{Normal}\left(0, \tilde{a}^2\right) & \text{if } \tilde{a} \neq 0; \\ \delta_0 & \text{otherwise.} \end{cases} \tag{3.30}$$

Observe that,

$$\frac{W_n - a(\log n)\mathbf{v}}{b(\log n)} = \frac{b(1+\tau_n)}{b(\log n)} \frac{W_n - a(1+\tau_n)\mathbf{v}}{b(1+\tau_n)} + \frac{\sqrt{\log n}}{b(\log n)} \frac{a(1+\tau_n) - a(\log n)}{\sqrt{\log n}}\mathbf{v}. \tag{3.31}$$

First note that by the Converging Together Lemma, we get that for the first term of the equation above

$$\text{dist}\left(\frac{W_n - a(1+\tau_n)\mathbf{v}}{b(\log n)} \,\Big|\, \mathcal{G}_n\right) \xrightarrow{p} \Lambda \text{ in } \mathcal{P}\left(\mathbb{R}^d\right). \tag{3.32}$$

Further, for the second term, again by the Converging Together Lemma, and the equation (3.30) we have

$$\frac{a(1+\tau_n) - a(\log n)}{b(\log n)} \Rightarrow \begin{cases} \text{Normal}\left(0, \tilde{a}^2\tilde{b}^2\right) & \text{if } \tilde{a}\tilde{b} \neq 0; \\ \delta_0 & \text{otherwise.} \end{cases} \tag{3.33}$$

Now, let

$$A_n := \frac{W_n - a(1+\tau_n)\mathbf{v}}{b(\log n)} \text{ and } B_n := \frac{a(1+\tau_n) - a(\log n)}{b(\log n)}.$$

For $\mathbf{t} \in \mathbb{R}^d$ define $\phi^{(r)}_{A_n}(\mathbf{t}) := \mathbf{E}\left[\exp\left(\mathrm{i}A_n\mathbf{t}^T\right) \,\Big|\, \mathcal{G}_n\right]$ and for $t \in \mathbb{R}$ define $\phi_{B_n}(t) := \mathbf{E}\left[\exp\left(\mathrm{i}tB_n\right)\right]$. Note these are the conditional characteristic function of A_n given $\mathcal{G}_n$ and the (unconditional or marginal) characteristic function of B_n. Finally, for $\mathbf{t} \in \mathbb{R}^d$, let $\phi_n(\mathbf{t}) := \mathbf{E}\left[\exp\left(\mathrm{i}\frac{W_n - a(\log n)\mathbf{v}}{b(\log n)}\mathbf{t}^T\right)\right]$, be the characteristic function of $\frac{W_n - a(\log n)\mathbf{v}}{b(\log n)}$. Then

$$\begin{aligned} \phi_n(\mathbf{t}) &= \mathbf{E}\left[\mathbf{E}\left[\exp\left(\mathrm{i}A_n\mathbf{t}^T\right) \,\Big|\, \mathcal{G}_n, \mathcal{T}_{n-1}\right]\right] \\ &= \mathbf{E}\left[\phi^{(r)}_{A_n}(\mathbf{t}) \sum_{m=-1}^{n-1} \exp\left(\mathrm{i}\frac{a(1+|m|_{\mathcal{T}}) - a(\log n)}{b(\log n)}\mathbf{v}\mathbf{t}^T\right)\right]. \end{aligned} \tag{3.34}$$

The last equality follows because of the observation (3.3). Now by equation (3.32) we get

$$\phi^{(r)}_{A_n}(\mathbf{t}) \xrightarrow{p} \int_{\mathbb{R}^d} e^{\mathrm{i}\mathbf{x}\mathbf{t}^T}\,\Lambda(d\mathbf{x}). \tag{3.35}$$

But since all quantities are bounded (in fact, bounded by 1), so we conclude that

$$\mathbf{E}\left[\left|\phi^{(r)}_{A_n}(\mathbf{t}) - \int_{\mathbb{R}^d} e^{\mathrm{i}\mathbf{x}\mathbf{t}^T}\,\Lambda(d\mathbf{x})\right|\right] \longrightarrow 0.$$

Further,

$$\begin{aligned}\phi_{B_n}\left(\mathbf{vt}^T\right) &= \mathbf{E}\left[\exp\left(\mathrm{i}B_n\mathbf{vt}^T\right)\right]\\ &= \mathbf{E}\left[\mathbf{E}\left[\exp\left(\mathrm{i}B_n\mathbf{vt}^T\right)\,\middle|\,\mathcal{T}_{n-1}\right]\right]\\ &= \mathbf{E}\left[\sum_{m=-1}^{n-1}\exp\left(\mathrm{i}\frac{a\left(1+|m|_{\mathcal{T}}\right)-a\left(\log n\right)}{b\left(\log n\right)}\mathbf{vt}^T\right)\right]\\ &\longrightarrow e^{-\tilde{a}^2\tilde{b}^2\frac{\left(\mathbf{vt}^T\right)^2}{2}},\end{aligned}\tag{3.36}$$

where the convergence follows from the equation (3.33). Finally, using equations (3.34), (3.35) and (3.36) we conclude that

$$\phi_n\left(\mathbf{t}\right)\longrightarrow e^{-\tilde{a}^2\tilde{b}^2\frac{\left(\mathbf{vt}^T\right)^2}{2}}\cdot\int_{\mathbb{R}^d}e^{\mathrm{i}\mathbf{xt}^T}\,\Lambda\left(d\mathbf{x}\right)\ \ \forall\,\mathbf{t}\in\mathbb{R}^d.$$

This proves the conclusion of Part (c). $\square$

LEMMA 3.6. *Let S_n^H be the set of vertices of the random recursive tree $\mathcal{T}_n$ up to depth $1\leq H\leq(n+1)$. Then*

$$\mathbf{E}\left[\left|S_n^H\right|\right]=\mathcal{O}\left(\left(\log\left(n+2\right)\right)^H\right),\tag{3.37}$$

where $\left|S_n^H\right|$ denotes the cardinality of the set S_n^H.

PROOF. We prove this by induction on H. First let us fix $H=1$, which is the base case of induction. Observe that any vertex is at depth 1, if and only if, it attaches itself to the root and this happens with probability $\frac{1}{n+1}$ for the n-th vertex.

This implies that

$$\left|S_{n+1}^1\right|=\begin{cases}\left|S_n^1\right|+1, & \text{w.p. } \frac{1}{n+2},\\ \left|S_n^1\right|, & \text{w.p. } (1-\frac{1}{n+2}).\end{cases}\tag{3.38}$$

Therefore, it follows that

$$\mathbf{E}\left[\left|S_{n+1}^1\right|\,\middle|\,\mathcal{T}_n\right]=(\left|S_n^1\right|+1)\frac{1}{n+2}+\left|S_n^1\right|(1-\frac{1}{n+2})=\left|S_n^1\right|+\frac{1}{n+2}.\tag{3.39}$$

The rest of the proof for $H=1$ is immediate since

$$\mathbf{E}\left[\left|S_{n+1}^1\right|\right]=\sum_{j=0}^{n+1}\frac{1}{j+1}=\mathcal{O}\left(\log(n+2)\right).$$

Let us assume that the result holds for $H-1$. Observe that as in (3.38), given $\mathcal{T}_n$, we have

$$\left|S_{n+1}^H\right|=\begin{cases}\left|S_n^H\right|+1, & \text{w.p. } \frac{\left|S_n^{H-1}\right|}{n+2},\\ \left|S_n^H\right|, & \text{w.p. } (1-\frac{\left|S_n^{H-1}\right|}{n+2}).\end{cases}\tag{3.40}$$

This implies by arguments similar to (3.39), we obtain for general H

$$\mathbf{E}\left[\left|S_{n+1}^H\right|\,\middle|\,\mathcal{T}_n\right]=\left|S_n^H\right|+\frac{\left|S_n^{H-1}\right|}{n+2}.\tag{3.41}$$

Hence,

$$\mathbf{E}\left[\left|S_{n+1}^H\right|\right]=\sum_{j=H-1}^{n+1}\frac{\left|S_j^{H-1}\right|}{j+1}=\mathcal{O}\left(\left(\log(n+2)\right)^H\right).$$

$\square$

3.3. *Asymptotic of Empirical Law of the Branching Markov Chain.* The following theorem can be proved using the same techniques as in the proof of the Theorem 3.4. In fact, the arguments are almost same but for sake of completeness, we provide the details.

THEOREM 3.7. *Suppose that the assumption **(A)** holds. Let $\mathcal{E}_n^{cs(r)}$ be a centered and scaled version of $\mathcal{E}_n$ with (possibly random) centering by $a(\tau_n)\mathbf{v}$ and scaling by $b(\tau_n)$. Then under the assumptions stated in the parts (a), (b) & (c) of the Theorem 3.4 and as $n \to \infty$,*

$$\mathcal{E}_n^{cs(r)} \xrightarrow{p} \Lambda, \tag{3.42}$$

where the above convergence is in $\mathcal{P}(S)$ under the conditions given in (a) below, otherwise in $\mathcal{P}(\mathbb{R}^d)$. Moreover, let $\mathcal{E}_n^{cs}$ be the centered and scaled version of $\mathcal{E}_n$ with (non-random) centering by $a(\log n)\mathbf{v}$ and scaling by $b(\log n)$, then

(a) *If $a = 0$ and $b = 1$, then*

$$\mathcal{E}_n^{cs} = \mathcal{E}_n \xrightarrow{p} \Lambda \text{ in } \mathcal{P}(S). \tag{3.43}$$

(b) *If the conditions of part (b) of the Theorem 3.4 hold, then*

$$\mathcal{E}_n^{cs} \xrightarrow{p} \Lambda \text{ in } \mathcal{P}\left(\mathbb{R}^d\right). \tag{3.44}$$

(c) *If the conditions of part (c) of the Theorem 3.4 hold, then*

$$\mathcal{E}_n^{cs} \xrightarrow{p} \Xi \text{ in } \mathcal{P}\left(\mathbb{R}^d\right), \tag{3.45}$$

where Ξ is Λ if $\tilde{a} = 0$ or $\tilde{b} = 0$, otherwise, it is given by the convolution of Λ and Normal $\left(0, \tilde{a}^2\tilde{b}^2\right)\mathbf{v}$.

PROOF. Let

$$\hat{\mathcal{E}}_n := \frac{1}{n+1}\sum_{k=-1}^{n-1} \delta_{W_k} = \frac{1}{n+1}\delta_\Delta + \frac{n}{n+1}\mathcal{E}_n, \tag{3.46}$$

which is the *empirical measure* of the variables $(W_k)_{k=-1}^{n-1}$. From (3.5) it then follows that

$$Q_n = \hat{\mathcal{E}}_n \hat{R}. \tag{3.47}$$

Informally, we can think of Q_n as "one more step taken" by a $\hat{R}$-chain when its current law is $\hat{\mathcal{E}}_n$. Note that for any $A \in \mathcal{B}_{\mathbb{R}^d}$,

$$\left|\hat{\mathcal{E}}_n(A) - \mathcal{E}_n(A)\right| \le \frac{1}{n+1}. \tag{3.48}$$

Thus almost surely,

$$\| \hat{\mathcal{E}}_n - \mathcal{E}_n \|_{\mathrm{TV}} \longrightarrow 0. \tag{3.49}$$

So it is enough to prove the asymptotic results in equations (3.42), (3.43), (3.44) and (3.45) for $\hat{\mathcal{E}}_n$ instead of $\mathcal{E}_n$.

Now, we recall that from the equation (3.47), our basic intuition that, Q_n is "one more step taken" by a $\hat{R}$-chain when its current law is $\hat{\mathcal{E}}_n$. In other words, $\hat{\mathcal{E}}_n$ is "one less step taken". This intuition works for making the proofs of the various parts of Theorem 3.4 work as well for proving the asymptotic of $\hat{\mathcal{E}}_n$.

We will start by proving that under the assumption **(A)**, the equation (3.42) holds.

Proof of Equation (3.42). STEP I: To show that $\mathbf{E}\left[\hat{\mathcal{E}}_n^{cs(r)}(\cdot)\right] \longrightarrow \Lambda(\cdot)$ in $\mathcal{P}\left(\mathbb{R}^d\right)$:

Recall, ρ is a metric on $\mathcal{P}(S)$, which metrizes the weak convergence topology on it. Also recall, L_n denotes the distribution of $\frac{X_n - a(n)}{b(n)}$. For $\epsilon > 0$, let H and N (which may depend on $\epsilon > 0$) be chosen as in the proof of Theorem 3.4.

Recall that, D_n is the vertex in the random recursive tree $\mathcal{T}_{n-1}$ at which the n-th vertex joins. So from definition, $\hat{\mathcal{E}}_n = \mathbf{E}\left[\delta_{W_{D_n}} \middle| \mathcal{G}_n\right]$. Now let $\tilde{\mathcal{E}}_n := \mathbf{E}\left[\delta_{W_{D_n}} \middle| \mathcal{T}_n\right]$. Thus, $\mathbf{E}\left[\hat{\mathcal{E}}_n^{cs(r)}(\cdot)\right] = \mathbf{E}\left[\tilde{\mathcal{E}}_n^{cs}(\cdot)\right]$, where $\mathbf{E}\left[\tilde{\mathcal{E}}_n^{cs}(\cdot)\right]$ is a centered and scaled version of $\tilde{\mathcal{E}}_n$ with centering by $a(\tau_n)\mathbf{v}$ and scaling by $b(\tau_n)$. So it is enough to show that $\mathbf{E}\left[\tilde{\mathcal{E}}_n^{cs(r)}(\cdot)\right] \longrightarrow \Lambda(\cdot)$ in $\mathcal{P}\left(\mathbb{R}^d\right)$.

Now, notice that from definition $|D_n|_{\mathcal{T}} = \tau_n$, thus, given $\mathcal{T}_n$,

$$\rho\left(\tilde{\mathcal{E}}_n^{cs(r)}, \Lambda\right) \stackrel{d}{=} \sum_{j=0}^{n} \mathbf{1}_{[\tau_n = j]}\, \rho\left(L_j, \Lambda\right). \tag{3.50}$$

So with the same quantity S_n^H as defined in the proof of Theorem 3.4 and using Lemma 3.6, we get

$$\mathbf{P}\left(\rho\left(\tilde{\mathcal{E}}_n^{cs(r)}, \Lambda\right) > \epsilon\right) \leq \mathbf{P}\left(D_n \in S_n^H\right) = \frac{\mathbf{E}\left[\left|S_n^H\right|\right]}{n+2} < \epsilon, \tag{3.51}$$

In other words,

$$\tilde{\mathcal{E}}_n^{cs(r)}(\cdot) \xrightarrow{p} \Lambda(\cdot) \;\Rightarrow\; \mathbf{E}\left[\tilde{\mathcal{E}}_n^{cs(r)}(\cdot)\right] \longrightarrow \Lambda(\cdot). \tag{3.52}$$

This completes the proof of the STEP I.

Let $C(\Lambda) \subseteq \mathbb{R}^d$ be as defined in the proof of Theorem 3.4

STEP II: To show that $\mathbf{Var}\left(\hat{\mathcal{E}}_n^{cs(r)}(-\infty, \mathbf{x}]\right) \longrightarrow 0 \;\; \forall \mathbf{x} \in C(\Lambda)$:

Fix $\mathbf{x} \in C(\Lambda) \subseteq \mathbb{R}^d$. Recall, that from definition

$$\hat{\mathcal{E}}_n^{cs(r)}(-\infty, \mathbf{x}] = \frac{1}{n+1} \sum_{k=-1}^{n-1} \mathbf{1}\left((W_k \in (-\infty, a(1 + |k|_{\mathcal{T}})\mathbf{v} + b(1 + |k|_{\mathcal{T}})\mathbf{x}]\right)$$

Now, as in the proof of Theorem 3.4, let (u_n, v_n) be two vertices chosen uniformly at random from the set of vertices of $\mathcal{T}_{n-1}$. Then

$$\begin{aligned}
&\mathbf{Var}\left(\hat{\mathcal{E}}_n^{cs(r)}((-\infty, \mathbf{x}])\right) \\
&= \frac{1}{(n+1)^2} \sum_{m,j=-1}^{n-1} \mathbf{Cov}\left(\mathbf{1}\left(W_m \in (-\infty, a(|m|_{\mathcal{T}})\mathbf{v} + b(|m|_{\mathcal{T}})\mathbf{x}]\right),\right. \\
&\qquad\qquad \left.\mathbf{1}\left(W_j \in (-\infty, a(|j|_{\mathcal{T}})\mathbf{v} + b(|j|_{\mathcal{T}})\mathbf{x}]\right)\right) \\
&= \mathbf{Cov}\left(\mathbf{1}\left(\frac{W_{u_n} - a(|u_n|_{\mathcal{T}})\mathbf{v}}{b(|u_n|_{\mathcal{T}})} \leq \mathbf{x}\right), \mathbf{1}\left(\frac{W_{v_n} - a(|v_n|_{\mathcal{T}})\mathbf{v}}{b(|v_n|_{\mathcal{T}})} \leq \mathbf{x}\right)\right)
\end{aligned}$$

$$
\begin{aligned}
&= \mathbf{P}\left(\frac{W_{u_n} - a\left(|u_n|_{\mathcal{T}}\right)\mathbf{v}}{b\left(|u_n|_{\mathcal{T}}\right)} \le \mathbf{x} \text{ and } \frac{W_{v_n} - a\left(|v_n|_{\mathcal{T}}\right)\mathbf{v}}{b\left(|v_n|_{\mathcal{T}}\right)} \le \mathbf{x}\right) \\
&\qquad - \left(\mathbf{P}\left(\frac{W_{u_n} - a\left(|u_n|_{\mathcal{T}}\right)\mathbf{v}}{b\left(|u_n|_{\mathcal{T}}\right)} \le \mathbf{x}\right)\right)^2
\end{aligned} \tag{3.53}
$$

Now the second term in the above equation is exactly same as

$$
\frac{1}{n+1}\sum_{k=-1}^{n-1}\left(\mathbf{P}\left(\frac{W_k - a\left(|k|_{\mathcal{T}}\right)\mathbf{v}}{b\left(|k|_{\mathcal{T}}\right)} \le \mathbf{x}\right)\right)^2.
$$

But, by Theorem 2.8 and assumption (A), we get $\frac{W_k - a(|k|_{\mathcal{T}})\mathbf{v}}{b(|k|_{\mathcal{T}})} \Rightarrow \Lambda$. Thus, the second term in (3.53) converges to $\Lambda^2\left((-\infty, \mathbf{x}]\right)$. So it is enough to show that the first term also converges to $\Lambda^2\left((-\infty, \mathbf{x}]\right)$.

Now as in the proof of Theorem 3.4, let ξ_n be the *least common ancestor* of the vertices u_n and v_n in $\mathcal{T}_{n-1}$. Then

$$
\begin{aligned}
&\mathbf{P}\left(\frac{W_{u_n} - a\left(|u_n|_{\mathcal{T}}\right)\mathbf{v}}{b\left(|u_n|_{\mathcal{T}}\right)} \le \mathbf{x} \text{ and } \frac{W_{v_n} - a\left(|v_n|_{\mathcal{T}}\right)\mathbf{v}}{b\left(|v_n|_{\mathcal{T}}\right)} \le \mathbf{x}\right) \\
&= \mathbf{E}\left[\mathbf{P}\left(\frac{W_{u_n} - a\left(|u_n|_{\mathcal{T}}\right)\mathbf{v}}{b\left(|u_n|_{\mathcal{T}}\right)} \le \mathbf{x} \text{ and } \frac{W_{v_n} - a\left(|v_n|_{\mathcal{T}}\right)\mathbf{v}}{b\left(|v_n|_{\mathcal{T}}\right)} \le \mathbf{x} \,\middle|\, \xi_n, W_{\xi_n}, \mathcal{T}\right)\right]
\end{aligned} \tag{3.54}
$$

Now, similar to the argument presented in the proof of Theorem 3.4, given $[|\xi_n|_{\mathcal{T}} = k, W_k, \mathcal{T}]$, the distribution of the variable W_{u_n} is same as the distribution of a Markov chain starting at δ_{W_k} with replacement kernel $\hat{R}$ and have taken a total of $(|u_n|_{\mathcal{T}} - k)$-many steps. Similar arguments follow for W_{v_n} and they are independent. So from (3.54) we get

$$
\begin{aligned}
&\mathbf{P}\left(W_{u_n} \le a\left(|u_n|_{\mathcal{T}}\right)\mathbf{v} + b\left(|u_n|_{\mathcal{T}}\right)\mathbf{x} \text{ and } W_{v_n} \le a\left(|v_n|_{\mathcal{T}}\right)\mathbf{v} + b\left(|v_n|_{\mathcal{T}}\right)\mathbf{x}\right) \\
&= \sum_{k=0}^{\infty}\mathbf{P}\left(|\xi_n|_{\mathcal{T}} = k\right)\int_S\int_S \hat{R}^k(\Delta, dt)\mathbf{E}\left[\hat{R}^{|u_n|_{\mathcal{T}}-k}\left(t, (-\infty, a\left(|u_n|_{\mathcal{T}}\right)\mathbf{v} + b\left(|u_n|_{\mathcal{T}}\right)\mathbf{x}]\right)\right. \\
&\qquad \left.\times \hat{R}^{|v_n|_{\mathcal{T}}-k}\left(t, (-\infty, a\left(|v_n|_{\mathcal{T}}\right)\mathbf{v} + b\left(|v_n|_{\mathcal{T}}\right)\mathbf{x}]\right)\right]
\end{aligned} \tag{3.55}
$$

Now, following same argument as in Theorem 3.4 leading to equation (3.21), we conclude that $\epsilon > 0$ there exists $N \ge -1$, such that,

$$
\begin{aligned}
&\left|\mathbf{P}\left(\frac{W_{u_n} - a\left(|u_n|_{\mathcal{T}}\right)\mathbf{v}}{b\left(|u_n|_{\mathcal{T}}\right)} \le \mathbf{x} \text{ and } \frac{W_{v_n} - a\left(|v_n|_{\mathcal{T}}\right)\mathbf{v}}{b\left(|v_n|_{\mathcal{T}}\right)} \le \mathbf{x}\right)\right. \\
&- \sum_{k=0}^{N}\mathbf{P}\left(|\xi_n|_{\mathcal{T}} = k\right)\int_S\int_S \hat{R}^k(\Delta, dt)\mathbf{E}\left[\hat{R}^{|u_n|_{\mathcal{T}}-k}\left(t, (-\infty, a\left(|u_n|_{\mathcal{T}}\right)\mathbf{v} + b\left(|u_n|_{\mathcal{T}}\right)\mathbf{x}]\right)\right. \\
&\qquad \left.\left.\times \hat{R}^{|v_n|_{\mathcal{T}}-k}\left(t, (-\infty, a\left(|v_n|_{\mathcal{T}}\right)\mathbf{v} + b\left(|v_n|_{\mathcal{T}}\right)\mathbf{x}]\right)\right]\right| \le \epsilon
\end{aligned} \tag{3.56}
$$

Now an argument similar to that in the proof of Theorem 3.4 leading to equation 3.22 will show

$$
\mathbf{P}\left(\frac{W_{u_n} - a\left(|u_n|_{\mathcal{T}}\right)\mathbf{v}}{b\left(|u_n|_{\mathcal{T}}\right)} \le \mathbf{x} \text{ and } \frac{W_{v_n} - a\left(|v_n|_{\mathcal{T}}\right)\mathbf{v}}{b\left(|v_n|_{\mathcal{T}}\right)} \le \mathbf{x}\right) \longrightarrow \Lambda^2\left((-\infty, \mathbf{x}]\right). \tag{3.57}
$$

This completes the proof of the Step II.

STEP III: To show that equation (3.42) holds:

Now using the STEPS I & II we get that

$$\hat{\mathcal{E}}_n^{cs(r)}\left(-\infty,\mathbf{x}\right] \xrightarrow{p} \Lambda\left(-\infty,\mathbf{x}\right] \ \ \forall \mathbf{x} \in C\left(\Lambda\right). \tag{3.58}$$

The rest follows from a diagonal argument as described in the Step III of the proof of the Theorem (3.4)

Rest of the proof, that is the proofs of parts (a), (b) & (c) can be written almost verbatim by looking into the proofs of the parts (a), (b) & (c) of Theorem 3.4. □

3.4. *Asymptotic of the Random Configuration of the Urn.* Let $\mathcal{F}_n$ be the σ-algebra $\sigma\left(Z_0, Z_1, \cdots, Z_{n-1}; U_0\right)$, $n \geq 0$. Let P_n be a version of the regular conditional distribution of Z_n given $\mathcal{F}_n$. Note that by construction $P_n = \frac{U_n}{n+1}$ almost surely. The following result is an immediate corollary of the Theorem 2.4 and Theorem 3.4.

THEOREM 3.8. *Suppose that the assumption **(A)** holds. Let P_n^{cs} is the conditional distribution of $\frac{Z_n - a(\log n)\mathbf{v}}{b(\log n)}$ given $\mathcal{F}_n$, that is, a scaled and centered version of P_n with centering by $a\left(\log n\right)\mathbf{v}$ and scaling by $b\left(\log n\right)$, then*

(a) *If $a = 0$ and $b = 1$, then*

$$P_n^{cs} = P_n \xrightarrow{p} \Lambda \text{ in } \mathcal{P}\left(S\right). \tag{3.59}$$

(b) *If the conditions of part (b) of the Theorem 3.4 hold, then*

$$P_n^{cs} \xrightarrow{p} \Lambda \text{ in } \mathcal{P}\left(\mathbb{R}^d\right). \tag{3.60}$$

(c) *If the conditions of the part (c) of the Theorem 3.4 hold, then*

$$P_n^{cs} \xrightarrow{p} \Xi \text{ in } \mathcal{P}\left(\mathbb{R}^d\right), \tag{3.61}$$

where Ξ is Λ if $\tilde{a} = 0$ or $\tilde{b} = 0$, otherwise, it is given by the convolution of Λ and Normal $\left(0, \tilde{a}^2\tilde{b}^2\right)\mathbf{v}$.

3.5. *Asymptotic of the Expected Configuration of the Urn.* Recall that $\mathbf{E}\left[P_n\right] = \frac{\mathbf{E}[U_n]}{n+1}$ is the marginal distribution of Z_n. The following result is an immediate corollary of the Theorem 3.8.

THEOREM 3.9. *Suppose that the assumption **(A)** holds, then*

(a) *If $a = 0$ and $b = 1$, then*

$$Z_n \Rightarrow \Lambda. \tag{3.62}$$

(b) *If the conditions of part (b) of the Theorem 3.4 hold, then*

$$\frac{Z_n}{b\left(\log n\right)} \Rightarrow \Lambda, \tag{3.63}$$

(c) *If the conditions of part (c) of the Theorem 3.4 hold, then*

$$\frac{Z_n - a\left(\log n\right)\mathbf{v}}{b\left(\log n\right)} \Rightarrow \Xi, \tag{3.64}$$

where Ξ is Λ if $\tilde{a} = 0$ or $\tilde{b} = 0$, otherwise, it is given by the convolution of Λ and Normal $\left(0, \tilde{a}^2\tilde{b}^2\right)\mathbf{v}$.

PROOF. The result follows from the Theorem 3.8 by taking expectation and noting the fact that centering and scaling are non-random in all cases. □

3.6. *Asymptotic of the Color Count Statistics.* The following result follows immediately using Corollary 2.7 and Theorem 3.7.

THEOREM 3.10. *Suppose that the assumption* **(A)** *holds. Let $\mathcal{N}_n^{cs}$ be a scaled and centered version of $\mathcal{N}_n$ with centering by $a(\log n)\,\mathbf{v}$ and scaling by $b(\log n)$, then*

(a) *If $a=0$ and $b=1$, then*

$$\frac{\mathcal{N}_n^{cs}}{n} = \frac{\mathcal{N}_n}{n} \xrightarrow{p} \Lambda \text{ in } \mathcal{P}(S). \tag{3.65}$$

(b) *If the conditions of part (b) of the Theorem 3.4 hold, then*

$$\frac{\mathcal{N}_n^{cs}}{n} \xrightarrow{p} \Lambda \text{ in } \mathcal{P}\left(\mathbb{R}^d\right). \tag{3.66}$$

(c) *If the conditions of the part (c) of the Theorem 3.4 hold, then*

$$\frac{\mathcal{N}_n^{cs}}{n} \xrightarrow{p} \Xi \text{ in } \mathcal{P}\left(\mathbb{R}^d\right), \tag{3.67}$$

where Ξ is Λ if $\tilde{a}=0$ or $\tilde{b}=0$, otherwise, it is given by the convolution of Λ and Normal $\left(0,\tilde{a}^2\tilde{b}^2\right)\mathbf{v}$.

4. Applications in Various Urn Models. In this section we discuss several applications of the representation theorems (Theorem 2.4 and Theorem 2.8) for deriving results on various urn schemes. Essentially all the results stated here are proved using the two general asymptotic results, namely, Theorem 3.8 and Theorem 3.9, given in the previous section.

4.1. *S is Countable.*

4.1.1. *R is Ergodic.* Suppose the indexing set of colors S is either finite or countably infinite and we endow S with the $sigma$-algebra $\mathcal{S}$, which is the power set $\wp(S)$. In this case, we can view the Markov transition kernel R as a matrix and it is then called the *replacement matrix*. For S finite, it is the classical case. If we assume that R is ergodic, that is, assumption **(A)** holds with $a=0$ and $b=1$, then from Theorems 3.8 (a) and 3.10 (a) we get the following result.

THEOREM 4.1. *Suppose S is countable, $\mathcal{S}=\wp(S)$, R is ergodic with stationary distribution π on S. Then as $n\to\infty$,*

$$\frac{U_n}{n+1} \xrightarrow{p} \pi \text{ in } \mathcal{P}(S). \tag{4.1}$$

In particular,

$$\frac{\mathbf{E}[U_n]}{n+1} \xrightarrow{w} \pi, \tag{4.2}$$

as $n\to\infty$. Further,

$$\frac{\mathcal{N}_n}{n} \xrightarrow{p} \pi \text{ in } \mathcal{P}(S). \tag{4.3}$$

If S is finite then using either matrix algebra techniques or multi-type branching process techniques, it is known [27, 28] that stronger result holds. In fact, under even weaker assumption of only *irreducibility* of the chain, the convergence in probability in (4.1) can be replaced by almost sure convergence. We believe that in general for S countable, under ergodicity assumption almost sure convergence should hold. Here we note that as soon as S is infinite, the classical techniques such as matrix algebra methods using *Perron-Frobenius theory* of matrices with positive entries [42] and *Jordan Decomposition* of finite dimensional matrices [17], or martingale approach using embedding to multi-type branching processes, which have been extensively used in classical urn model literature [1, 27, 28, 29, 3, 10, 18]; fails to derive any result. We are hopeful that our novel and fairly probabilistic approach, namely, the *Grand and Marginal representation Theorems* should yield the classical result. Unfortunately, we have been unable to derive it so far.

4.1.2. *R is Block Diagonal.* Similar to the previous section, suppose the indexing set of colors S is either finite or countably infinite and we endow S with the $sigma$-algebra $\mathcal{S}$, which is the power set $\wp(S)$. As in the previous case, we view the Markov transition kernel R as a matrix. Suppose the indexing set of colors can be partitioned as $S = \underset{i \in \mathcal{I}}{\cup} C_i$, where $\mathcal{I}$ is a countable set, and C_i is countable for all $i \in \mathcal{I}$. We endow $\mathcal{I}$ with its power set as a σ-algebra on it and each C_i is also endowed with its power set as the σ-algebra on it.

Now let, $\phi : S \to \mathcal{I}$ be the "projection" map, which maps $s \mapsto i$, where i is the unique element of $\mathcal{I}$, such that, $s \in C_i$.

Now, suppose for every $i \in \mathcal{I}$ and $s \in C_i$, the kernel $R(s, \cdot)$ is a probability measure supported only on C_i, that is,

$$R(s, C_i) = \begin{cases} 1 \text{ if } s \in C_i \\ 0 \text{ otherwise.} \end{cases} \tag{4.4}$$

As each C_i is countable, R on C_i can be realized as a (possibly infinite) matrix R_{ii} indexed by the colors in C_i.

Note that if S is finite then R is essentially a reducible matrix with diagonal blocks, which can be presented as

$$R = \begin{pmatrix} R_{11} & 0 & 0 & \cdots & 0 \\ 0 & R_{22} & 0 & \cdots & 0 \\ 0 & 0 & R_{33} & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & R_{kk} \end{pmatrix},$$

We further assume that for all $i \in \mathcal{I}$, the kernel/replacement matrix, R_{ii} restricted to its "block" C_i, is ergodic with stationary distribution, π_i.

THEOREM 4.2. *Consider an urn model with colors indexed by a set S and replacement kernel R as in (4.4). Then for every initial configuration U_0, as $n \to \infty$,*

$$\frac{U_n}{n+1} \xrightarrow{p} \Pi \ \textit{in} \ \mathcal{P}(S), \tag{4.5}$$

where Π is a random probability measure on $(S, \mathcal{S})$ given by

$$\Pi(A) = \sum_{i \in \mathcal{I}} \pi_i (A \cap C_i)\, \nu_i, \ \ A \in \mathcal{S}, \tag{4.6}$$

and ν has Ferguson Distribution *on the countable set $\mathcal{I}$ with parameter $U_0 \circ \phi^{-1}$, and $\nu_i = \nu(\{i\})$, for $i \in \mathcal{I}$. Further,*

$$\frac{\mathcal{N}_n}{n} \xrightarrow{p} \Pi \ \textit{in} \ \mathcal{P}(S), \tag{4.7}$$

PROOF. Let us denote by $c_i := \sum_{v \in C_i} U_{0,v}$, and $T_{n,i} := \sum_{v \in C_i} U_{n,v}$ for each $i \geq 1$. It is easy to check that for each $i \geq 1$, the sequence $\left(\frac{T_{n,i}}{n+1}\right)_{n \geq 0}$ is non-negative a.s. convergent martingale, see [27] for details. Consider each C_i to be a super color for $i \geq 1$. Then $T_n := (T_{n,i})_{i \geq 1}$ corresponds to the configuration of a classical Pólya urn model, with initial configuration $T_0 = (c_i)_{i \geq 1}$. It is worthwhile to note here that $\frac{T_n}{n+1}$ is a probability measure on $\mathcal{I}$. Therefore, from [8], as $n \to \infty$

$$\frac{T_n}{n+1} \longrightarrow \nu \text{ a.s.},$$

where ν is a random measure on $\mathcal{I}$ having Ferguson Distribution with parameter $U_0 \circ \phi^{-1}$. In particular, almost surely, for any $i \in \mathcal{I}$,

$$\frac{T_{n,i}}{n+1} \longrightarrow \nu_i. \tag{4.8}$$

For any super color C_i, $i \geq 1$, define the sequence of random times

$$N_n(i) := \max\{k \leq n : Z_k \in C_i\} = \sum_{k=0}^{n-1} \mathbf{1}_{\{Z_k \in C_i\}}.$$

That is, $N_n(i)$ is the last time till n, a color has been chosen from the set C_i.

Then it is obvious that $T_{n,i} = T_{N_n(i),i} = c_i + N_n(i)$. From (4.8) we know that $N_n(i) \longrightarrow \infty$ a.s. as $n \to \infty$.

Denote by U_{n,C_i} the subvector of U_n corresponding to the color C_i. From [27], we know that $\left(U_{N_n(i),C_i}\right)$ is an urn model with initial configuration U_{0,C_i} and replacement matrix R_{ii}. Therefore, from Theorem 4.1, we know that

$$\frac{U_{N_n(i)}}{T_{N_n(i),i}} \xrightarrow{p} \pi_i, \text{ as } n \to \infty.$$

This implies that

$$\frac{U_{n,C_i}}{n+1} = \frac{U_{N_n(i),C_i}}{n+1} = \frac{U_{N_n(i),C_i}}{T_{N_n(i),i}} \frac{T_{N_n(i),i}}{n+1} = \frac{U_{N_{n,i},C_i}}{T_{N_n(i),i}} \frac{T_{n,i}}{n+1} \xrightarrow{p} \nu_i \, \pi_i \tag{4.9}$$

where ν as in (4.8). This completes the proof. Finally, same argument proves (4.7). ☐

4.2. *Urn Models Associated with Random Walks on* $\mathbb{R}^d$. It this section we consider urn models associated with random walks on *countable lattices* of $\mathbb{R}^d$. These models were first introduced in [6], where only bounded increment walks on $\mathbb{Z}^d$ were considered.

4.2.1. *Urn Models Associated with Random Walks on* $\mathbb{Z}^d$. Here we take $S = \mathbb{Z}^d$ for some $d \geq 1$, and $\mathcal{S}$ will be taken as the power set of $\mathbb{Z}^d$. The kernel R can be viewed as an infinite dimensional matrix index by the set of colors $\mathbb{Z}^d$, given by

$$R(u, v) = \mathbf{p}(v - u), \ u, v \in \mathbb{Z}^d, \tag{4.10}$$

where $\mathbf{p}$ is the distribution on $\mathbb{Z}^d$ of the independent increments of the walk.

Finite Variance Walks: Suppose $\mathbf{p}$ has finite second moment, leading to a random walk with finite variance. Following theorem is a generalization of the results derived in [6].

THEOREM 4.3. *Consider an infinite color urn model with colors indexed by $S = \mathbb{Z}^d$, and kernel R as given above. Suppose the starting configuration is U_0. Then there exist $\mu \in \mathbb{R}^d$ and a positive definite matrix $\Sigma_{d\times d}$, such that, if we define,*

$$P_n^{cs}(A) := \frac{U_n}{n+1}\left(\sqrt{\log n}A\Sigma^{1/2} + \mu \log n\right), \ \ A \in \mathcal{B}_{\mathbb{R}^d},$$

where

$$xA\Sigma^{1/2} := \{xy\Sigma^{1/2} \colon y \in A\},$$

then, as $n \to \infty$,

$$P_n^{cs} \xrightarrow{p} \Phi_d \text{ in } \mathcal{P}\left(\mathbb{R}^d\right). \tag{4.11}$$

In particular,

$$\frac{Z_n - \mu \log n}{\sqrt{\log n}} \Rightarrow \mathit{Normal}_d(0, \Sigma), \tag{4.12}$$

as $n \to \infty$. Also let,

$$\mathcal{N}_n^{cs}(A) := \mathcal{N}_n\left(\sqrt{\log n}A\Sigma^{1/2} + \mu \log n\right), \ \ A \in \mathcal{B}_{\mathbb{R}^d},$$

then

$$\frac{\mathcal{N}_n^{cs}}{n} \xrightarrow{p} \Phi_d \text{ in } \mathcal{P}\left(\mathbb{R}^d\right). \tag{4.13}$$

PROOF. Let X_n be the position of the random walk starting with $X_0 \sim U_0$ and independent increments with distribution given by $\mathbf{p}$. From the classical Central Limit Theorem [21], we get that

$$\frac{X_n - n\mu}{\sqrt{n}} \Rightarrow \text{Normal}_d\left(\mathbf{0}, \Sigma - \mu\mu^T\right), \tag{4.14}$$

where μ is the mean of the increment distribution and Σ is the second moment. Thus assumption **(A)** holds, with $\mathbf{v} = \mu$, $a(n) = n$, $b(n) = \sqrt{n}$ and $\Lambda = \text{Normal}_d\left(\mathbf{0}, \Sigma - \mu\mu^T\right)$.

We observe that the assumptions in Part (c) of Theorem 3.4 holds with $\tilde{a} = 1$ and $\tilde{b} = 1$. This completes the proof of (4.11), by observing that $\Xi = \text{Normal}_d(0, \Sigma)$.

Finally, (4.12) follows from Theorem 3.9(c) and 4.13 follows from Theorem 3.10. □

Infinite Variance Walks: In this section, we discuss some examples of infinite variance cases, which cannot be derived by the techniques developed in [6]. Let $\mathbf{p}$ be a *symmetric* distribution on $\mathbb{Z}$, which is in the domain of attraction of a symmetric α-stable distribution Λ, where $0 < \alpha < 2$. For sake of completeness we provide here the definition of symmetric α-stable distributions.

DEFINITION 4.4. A distribution Λ is said to have a *symmetric α-stable* distribution and usually denoted by $S\alpha S$, with $0 < \alpha < 2$, if for any $t \in \mathbb{R}$,

$$\mathbf{E}\left[e^{itV}\right] = \exp\left(-\sigma^\alpha |t|^\alpha\right), \tag{4.15}$$

for some $\sigma > 0$, where $V \sim \Lambda$.

Necessary and sufficient conditions, when $\mathbf{p}$ will be in the domain of attraction of such a Λ can be found from [22] (see Chapter XVII.5). In particular, when $\mathbf{p}$ is symmetric, it is enough to assume that there exists $0 < \alpha < 2$ and a slowly varying function $L(\cdot)$, such that,

$$\mathbf{P}(|Y| > n) = \frac{L(n)}{n^{\alpha}}, \ n \in \mathbb{N}, \tag{4.16}$$

where $Y \sim \mathbf{p}$ (see Theorem 2(c) of Chapter XVII.5 of [22]). In other words, if $(X_n)_{n\geq 0}$ are the positions of a random walk on $\mathbb{Z}$ starting at X_0 and with i.i.d. increments distributed according to the distribution $\mathbf{p}$ as given above, then

$$\frac{X_n}{b(n)} \Rightarrow \Lambda, \tag{4.17}$$

where $b(n) := n^{\frac{1}{\alpha}} h(n)$ for some slowly varying function h. Note that h can be explicitly computed from L (see equation (5.23) of Chapter XVII.5 of [22]). Following theorem is now an immediate consequence of Theorem 3.8(b), Theorem 3.9(b) and Theorem 3.10(b)

THEOREM 4.5. *Consider an infinite color urn model with colors indexed by $S = \mathbb{Z}$, and kernel R defined by (4.10), where $\mathbf{p}$ is as defined above. Suppose the starting configuration is U_0. Then there exists a slowly varying function h and a $S\alpha S$-distribution Λ, where α is as in (4.16), such that, if we define, P_n^{cs} as the conditional distribution of $\frac{Z_n}{(\log n)^{\frac{1}{\alpha}} h(\log n)}$ given $\mathcal{F}_n$, then, as $n \to \infty$,*

$$P_n^{cs} \xrightarrow{p} \Lambda \text{ in } \mathcal{P}\left(\mathbb{R}^d\right). \tag{4.18}$$

In particular,

$$\frac{Z_n}{(\log n)^{\frac{1}{\alpha}} h(\log n)} \Rightarrow \Lambda, \tag{4.19}$$

as $n \to \infty$. Further, let

$$\mathcal{N}_n^{cs}(A) := \mathcal{N}_n\left((\log n)^{\frac{1}{\alpha}} h(\log n) A\right), \ A \in \mathcal{B}_{\mathbb{R}^d},$$

where $\mathcal{N}_n$ is the corresponding color count statistics*, then*

$$\frac{\mathcal{N}_n^{cs}}{n} \xrightarrow{p} \Lambda \text{ in } \mathcal{P}\left(\mathbb{R}^d\right). \tag{4.20}$$

4.2.2. *Urn Models Associated with Periodic Random Walk on $\mathbb{R}^d$.* Let $\mathbb{H} = (V, E)$ be the hexagonal lattice in $\mathbb{R}^2$ [see Figure 1]. The vertex set can easily be partitioned into two non-empty subsets, $V = V_1 \cup V_2$, where V_1 and V_2 are disjoint, and the random walk on $\mathbb{H}$ is then a *periodic* chain. If the replacement kernel be denoted by R, then corresponding urn scheme with colors indexed by $\mathbb{H}$, is not covered by the earlier stated Theorem 4.3. For studying such cases, we consider the following slightly more general type of random walk on $\mathbb{R}^d$.

Let $\{Y_j(i), 1 \leq i \leq k, \ j \geq 1\}$ be a collection of independent random d-dimensional vectors, such that, for each fixed $i \in \{1, 2, \ldots k\}$, $(Y_j(i))_{j\geq 1}$ are i.i.d. We further assume that for each fixed $1 \leq i \leq k$, there exists a finite non-empty set $B_i \subset \mathbb{R}^d$, such that, $\mathbf{P}(Y_1(i) \in B_i) = 1$, and $B_i \cap B_j = \emptyset$, for any $1 \leq i, j \leq k$. That is, for each $i \in \{1, 2, \ldots, k\}$, we assume that the law of $Y_1(i)$ is bounded. For $1 \leq i \leq k$, we shall write

$$\begin{aligned} \mu(i) &:= \mathbf{E}\left[Y_1(i)\right], \\ \overline{\mu} &:= \tfrac{1}{k}\sum_{i=1}^{k}\mu(i), \\ \Sigma(i) &:= \mathbf{E}\left[Y_1^T(i)Y_1(i)\right]. \end{aligned} \tag{4.21}$$

We further assume that $\Sigma(i)$ is positive definite, for each $1 \leq i \leq k$. Let us denote by $\Sigma^{1/2}(i)$ the unique *positive definite square root* of $\Sigma(i)$. Note that, then $\overline{\Sigma} = \frac{1}{k}\sum_{i=1}^{k}\Sigma(i)$ is also *positive definite*. We denote by $\overline{\Sigma}^{1/2}$, the unique *positive definite square root* of $\overline{\Sigma}$.

For $n = mk + r$, where $m \in \mathbb{N} \cup \{0\}$, and $0 \leq r < k$, let

$$X_n = X_{mk} + Y_{m+1}(1) + Y_{m+1}(2) + \ldots + Y_{m+1}(r+1),$$

be the k-periodic random walk with increments $\{Y_j(i), 1 \leq i \leq k,\ j \geq 1\}$.

In the remainder of this subsection, we will consider an urn model $(U_n)_{n\geq 0}$, with colors indexed by $S = \mathbb{R}^d$, starting at some distribution U_0 on $\mathbb{R}^d$ and with a replacement kernel R associated with a periodic random walk with periodic increments as given above.

THEOREM 4.6. *Consider an infinite color urn model with colors indexed by $S = \mathbb{R}^d$, and kernel R as given above. Suppose the starting configuration is U_0. If we define,*

$$P_n^{cs}(A) := \frac{U_n}{n+1}\left(\sqrt{\log n}A\overline{\Sigma}^{1/2} + \overline{\mu}\log n\right),\ \ A \in \mathcal{B}_{\mathbb{R}^d},$$

where

$$xA\Sigma^{1/2} := \{xy\Sigma^{1/2} \colon y \in A\},$$

then, as $n \to \infty$,

$$P_n^{cs} \xrightarrow{p} \Phi_d \text{ in } \mathcal{P}\left(\mathbb{R}^d\right). \tag{4.22}$$

In particular,

$$\frac{Z_n - \overline{\mu}\log n}{\sqrt{\log n}} \Rightarrow Normal_d(0, \overline{\Sigma}), \tag{4.23}$$

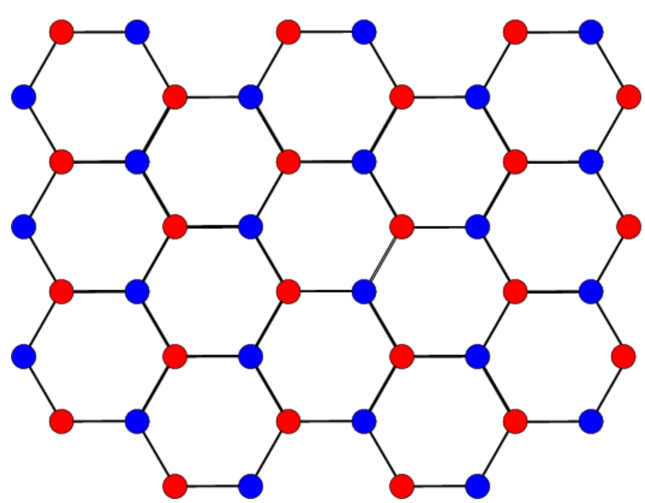

FIG 1. *Hexagonal Lattice*

as $n \to \infty$. Also let,

$$\mathcal{N}_n^{cs}(A) := \frac{U_n}{n+1}\left(\sqrt{\log n}A\overline{\Sigma}^{1/2} + \overline{\mu}\log n\right), \ \ A \in \mathcal{B}_{\mathbb{R}^d},$$

then

$$\frac{\mathcal{N}_n^{cs}}{n} \xrightarrow{p} \Phi_d \text{ in } \mathcal{P}\left(\mathbb{R}^d\right). \tag{4.24}$$

PROOF. We first note that by Theorem 3.8(c), Theorem 3.9(c) and Theorem 3.10, it is enough to show that

$$\frac{X_n - n\overline{\mu}}{\sqrt{n}} \Rightarrow \text{Normal}_d\left(0, \overline{D}\right), \tag{4.25}$$

where $\overline{D} = \frac{1}{k}\sum_{i=1}^k \mathbf{Var}\left(Y_1(i)\right)$. This follows from standard application of i.i.d. Central Limit Theorem [21]. □

As an application of the Theorem 4.6, we now consider our starting example of the random walk on hexagonal lattice. Let $\mathbb{H} = (V, E)$ be the hexagonal lattice in $\mathbb{R}^2$ [see Figure 1]. The vertex set $V = V_1 \cup V_2$, where V_1 and V_2 are disjoint. V_1 and V_2 are defined as follows:

$$V_{1,1} := \left\{1, \omega, \omega^2\right\}, \text{ where } \omega \text{ is a complex cube root of unity,}$$

and

$$V_{2,1} := \left\{v+1, v+\omega, v+\omega^2 \colon v \in V_{1,1}\right\}.$$

For any $n \geq 2$,

$$V_{1,n} := \left\{v-1, v-\omega, v-\omega^2 \colon v \in V_{2,n-1}\right\},$$

and

$$V_{2,n} = \left\{v+1, v+\omega, v+\omega^2 \colon v \in V_{1,n}\right\}.$$

Finally, $V_1 = \cup_{j\geq 1} V_{1,j}$ and $V_2 = \cup_{j\geq 1} V_{2,j}$. For any pair of vertices $v, w \in V$, we draw an edge between them, if and only if, either of the following two cases occur:

(i) $v \in V_1$ and $w \in V_2$ and $w = v + u$ for some $u \in \{1, \omega, \omega^2\}$, or
(ii) $v \in V_2$ and $w \in V_1$ and $w = v + u$ for some $u \in \{-1, -\omega, -\omega^2\}$.

To define the random walk on $\mathbb{H}$, let us consider $\{Y_j(i) \colon i = 1, 2, \ j \geq 1\}$ to be a sequence of independent random vectors such that $\left(Y_j(i)\right)_{j\geq 1}$ are i.i.d for every fixed $i = 1, 2$. Let $Y_1(1) \sim \text{Unif}\left\{1, \omega, \omega^2\right\}$, and $Y_1(2) \sim \text{Unif}\left\{-1, -\omega, -\omega^2\right\}$. One can now define a random walk on $\mathbb{H}$, with the increments $\{Y_j(i) \colon i = 1, 2, \ j \geq 1\}$. Needless to say, this random walk has period 2.

COROLLARY 4.7. Consider an infinite color urn model with colors indexed by $S = \mathbb{H}$, and kernel R as given above. Suppose the starting configuration is U_0. If we define,

$$P_n^{cs}(A) := \frac{U_n}{n+1}\left(2\sqrt{\log n}A\right), \ \ A \in \mathcal{B}_{\mathbb{R}^d},$$

then, as $n \to \infty$,

$$P_n^{cs} \xrightarrow{p} \Phi_2 \text{ in } \mathcal{P}\left(\mathbb{R}^2\right). \tag{4.26}$$

In particular,

$$\frac{2Z_n}{\sqrt{\log n}} \Rightarrow \Phi_2, \tag{4.27}$$

as $n \to \infty$

PROOF. First of all we note that, it is enough to show that $\overline{\Sigma} = \frac{1}{2}\mathbb{I}_2$.

Now, since $1+\omega+\omega^2 = 0$, so for the random walk on the hexagonal lattice, $\mu(1) = \mu(2) = 0$. Therefore $\overline{\mu} = 0$. Let

$$\Sigma(1) := \begin{pmatrix} \sigma_{1,1} \, \sigma_{1,2} \\ \sigma_{2,1} \, \sigma_{2,2} \end{pmatrix}$$

Writing $Y_1(1) := \left(Y_1^{(1)}(1), Y_1^{(2)}(1)\right)$, observe that

$$\sigma_{1,1} = \mathbf{E}\left[\left(Y_1^{(1)}(1)\right)^2\right] \text{ and } \sigma_{2,2} = \mathbf{E}\left[\left(Y_1^{(2)}(1)\right)^2\right].$$

Also,

$$\sigma_{1,2} = \sigma_{2,1} = \mathbf{E}\left[Y_1^{(1)}(1) Y_1^{(2)}(1)\right].$$

Writing $\omega = Re\,(\omega) + i Im\,(\omega)$, it is easy to see that

$$\sigma_{1,1} = \frac{1}{3}\left(1 + (Re\,(\omega))^2 + \left(Re\left(\omega^2\right)\right)^2\right).$$

Since $Re\,(\omega) = Re\left(\omega^2\right)$, therefore,

$$\sigma_{1,1} = \frac{1}{3}\left(1 + 2\,(Re\,(\omega))^2\right).$$

Since $\omega = \frac{1}{2} + i\frac{\sqrt{3}}{2}$, therefore, this implies $\sigma_{1,1} = \frac{1}{2}$. Similarly, since $Im\,(\omega) = -Im\left(\omega^2\right)$,

$$\sigma_{2,2} = \frac{1}{3}\left((Im\,(\omega))^2 + \left(Im\left(\omega^2\right)\right)^2\right) = \frac{2}{3}\,(Im\,(\omega))^2 = \frac{1}{2}.$$

Since, $Re\,(\omega) = Re\left(\omega^2\right)$, and $Im\,(\omega) = -Im\left(\omega^2\right)$,

$$\sigma_{1,2} = \sigma_{2,1} = \frac{1}{3}\left(Re\,(\omega)\, Im\,(\omega) + Re\left(\omega^2\right) Im\left(\omega^2\right)\right) = 0.$$

This proves that $\Sigma(1) = \frac{1}{2}\mathbb{I}_2$. Similar calculations show that $\Sigma(2) = \frac{1}{2}\mathbb{I}_2$. This implies that $\overline{\Sigma} = \frac{1}{2}\Sigma(1) + \frac{1}{2}\Sigma(2) = \frac{1}{2}\mathbb{I}_2$. This completes the proof. □

5. Application in Random Recursive Tree. As we have seen the *Grand Representation Theorem* (2.4) links the observed sequence of colors from an urn model with colors index by S, starting configuration U_0, a probability on $(S, \mathcal{S})$, and replacement kernel R to the corresponding *Branching Markov chain on the random recursive tree* as defined in the equations (2.1) and (2.7). In the previous section we saw several applications of the type that asymptotic properties of the urn model is derived by knowing the asymptotic properties of the associated Branching Markov chain on the random recursive tree. It is also possible that for certain choices of S, U_0 and R the asymptotic properties of the urn is well known and that then in turn one can get some non-trivial result about the associated Branching Markov chain on the random recursive tree. In particular, when R is *trivial* as a Markov kernel, that is, $R(s,\cdot) = \delta_s(\cdot)$ for all $s \in S$, then the urn scheme is the classical scheme of Pólya when S if finite [40], or the Blackwell and MacQueen Urn when S is a general Polish Space [8]. Following theorem about the sizes of the sub-trees rooted at the children of the root of a random recursive tree is an immediate consequence of the Grand Representation Theorem (2.4) and known facts about Pólya urn [40] and Blackwell and MacQueen Urn [8].

THEOREM 5.1. *Let $\mathcal{T}_n$ be the* random recursive tree *on $(n+2)$ vertices labeled by $\{-1;0,1,2,\ldots,n\}$, where the vertex labeled by -1 is considered as the* root. *Let N_n be the degree of the root vertex -1 and $S_1,S_2,\ldots,S_{N_n}$ be the sizes of the sub-trees rooted at the children of the root. For any distribution F on a Polish space $(S,\mathcal{S})$, if $\{\mathbf{X}_i\}_{i\geq 1}$ are* i.i.d. *variables taking values in S with distribution F and are independent of the random tree process $\{\mathcal{T}_n\}_{n\geq 1}$. Then*

$$\frac{1}{n+1}\sum_{i=1}^{N_n} S_i\,\delta_{\mathbf{X}_i} \xrightarrow{w} \mathcal{W}_F \text{ a.s.} \tag{5.1}$$

where $\mathcal{W}_F$ is a random probability on $(S,\mathcal{S})$ with Dirichlet(F) distribution. In particular, for any $0<p<1$, suppose $X_1,X_2,\ldots$ be an i.i.d. *sequence of Bernoulli(p) random variables, which are independent of the random tree process $\{\mathcal{T}_n\}_{n\geq 1}$. Then*

$$\frac{1}{n+1}\sum_{i=1}^{N_n} X_i S_i \longrightarrow W_p \text{ a.s.} \tag{5.2}$$

where $W_p \sim$ Beta$(p,1-p)$.

6. Urn Models with Random Replacement Scheme.

Given a measurable set $(S,\mathcal{S})$, let $\mathbb{P}$ be a probability on the set of all Markov transition kernels $R: S\times\mathcal{S}\to[0,1]$. A stochastic process $(U_n)_{n\geq 0}\subseteq\mathcal{M}(S)$ will be called an *urn model with random replacement scheme*, if U_0 is a probability measure on S, and

$$\mathbf{P}\left(Z_n\in ds\,\middle|\,U_n,U_{n-1},\cdots,U_0,\,R\right)=\frac{U_n(ds)}{n+1},\quad \mathbb{P}-\text{a.s.}, \tag{6.1}$$

and

$$U_{n+1}(A)=U_n(A)+R(Z_n,A),\quad A\in\mathcal{S}\quad \mathbb{P}-\text{a.s.} \tag{6.2}$$

Here we will consider Z_n as the color of the $(n+1)$-th selected ball. In other words, given a *random replacement scheme*, say R, the process $(U_n)_{n\geq 0}$ has the conditional law of an urn model with colors indexed by S, replacement scheme R and initial distribution U_0.

Given R, the law of $(U_n)_{n\geq 0}$, denoted by $\mathbf{P}\left(\cdot\middle|R\right)$ will be referred to as the *quenched law*. It is obvious that under the quenched law $(U_n)_{n\geq 0}$ is a Markov chain. The *annealed law* of the process is given by $\mathbb{E}\left[\mathbf{P}\left(\cdot\middle|R\right)\right]$, where the expectation is computed under the probability $\mathbb{P}$, determining the randomness in R. Note that, under the annealed law $(U_n)_{n\geq 0}$ need not be Markovian. We further note that under the annealed law the equations (1.3) and (1.4) hold.

It follows from Section 1.1 that along with $(U_n)_{n\geq 0}$ we can construct stochastic process $(X_n)_{n\geq 0}$ on S, such that, given the random replacement scheme R, under the quenched law, $\mathbf{P}\left(\cdot\middle|R\right)$, the sequence $(X_n)_{n\geq 0}$ is a Markov chain with transition kernel R and initial distribution U_0 and is independent of the urn process $(U_n)_{n\geq 0}$. Note that this process $(X_n)_{n\geq 0}$ need not remain Markov, nor can be independent of the urn process $(U_n)_{n\geq 0}$, under the annealed law. This stochastic process $(X_n)_{n\geq 0}$ will be called the *associated stochastic process*.

6.1. *Representation Theorems.*

Following result is an immediate consequence of the Theorem 2.4 for an urn model with random replacement scheme.

THEOREM 6.1. *Consider a random replacement scheme urn model $(U_n)_{n\geq 0}$ with colors indexed by a set S endowed with a σ-algebra $\mathcal{S}$. Let R be a random replacement kernel with distribution $\mathbb{P}$ and U_0 be the initial configuration. For $n\geq 0$, let Z_n be the random color of* *the* $(n+1)$*-th draw. Let* $(W_n)_{n\geq -1}$ *be such that, given* R*, it is a* branching Markov chain *on* $\mathcal{T}$ *as defined in* (2.7)*. Then under both the quenched and annealed laws*

$$(Z_n)_{n\geq 0} \stackrel{d}{=} (W_n)_{n\geq 0}\,. \tag{6.3}$$

The next result is a version of the Theorem 2.8 for the urn models with random replacement scheme.

THEOREM 6.2. *Consider a random replacement scheme urn model* $(U_n)_{n\geq 0}$ *with colors indexed by a set* S *endowed with a* σ*-algebra* $\mathcal{S}$*. Let* R *be a random replacement kernel with distribution* $\mathbb{P}$ *and* U_0 *be the initial configuration. For* $n \geq 0$*, let* Z_n *be the random color of the* $(n+1)$*-th draw. Let* $(X_n)_{n\geq 0}$ *be the associated stochastic process taking values on* S*, as defined above. Then there exists an increasing sequence of random indices* $(\tau_n)_{n\geq 0}$ *with* $\tau_0 = 0$*, which are independent of* $(X_n)_{n\geq 0}$ *under the quenched law; and are also independent of* R*, such that, for any* $n \geq 0$*,*

$$\mathbf{P}\left(Z_n \in A \,\middle|\, R\right) = \mathbf{P}\left(X_{\tau_n} \in A \,\middle|\, R\right), \quad \forall\ A \in \mathcal{S} \quad \mathbb{P}-a.s.. \tag{6.4}$$

In particular, under the annealed law, for any $n \geq 0$*,*

$$Z_n \stackrel{d}{=} X_{\tau_n}. \tag{6.5}$$

Moreover, the sequence of random indices $(\tau_n)_{n\geq 0}$ *satisfies equations* (2.2) *and* (2.3)*.*

PROOF. The equation (6.4) follows from the Theorem 2.8 and the equation (6.5) follows by taking expectation. □

6.2. *Applications.* In this section we give two examples to demonstrate that urn models on infinitely many colors and with random replacement scheme can have standard or non-standard limits depending on the behavior of the associated stochastic process. Both the examples are related to *random walks in random environment (RWRE)* on the one dimensional integer lattice $\mathbb{Z}$ and the colors will be indexed by $\mathbb{Z}$.

Consider $R = ((R(i,j)))_{i,j\in\mathbb{Z}}$, where we assume that there exists $0 < \delta < 1$, such that, for each $i \in \mathbb{Z}$, $\delta < R(i,i-1), R(i,i+1)$, and $R(i,i-1) + R(i,i+1) = 1$ a.s. and $(R(i,i+1))_{i\in\mathbb{Z}}$ are i.i.d. Note that in this case, the associated stochastic process $(X_n)_{n\geq 0}$ is a nearest neighbor RWRE on $\mathbb{Z}$ with i.i.d. environments, which is well studied in literature [45]. Recall that $\mathbb{P}$ and $\mathbb{E}$ denote respectively the distribution of R and expectation with respect to $\mathbb{P}$. We define

$$\rho_0 := \frac{R(0,-1)}{R(0,1)}.$$

and when $\mathbb{E}[\rho_0] < 1$, we define

$$v := \frac{1-\mathbb{E}[\rho_0]}{1+\mathbb{E}[\rho_0]}.$$

6.2.1. *Urn Associated with a Transient RWRE on* $\mathbb{Z}$*.*

THEOREM 6.3. *Consider an infinite color random replacement scheme urn model with colors indexed by* $S = \mathbb{Z}$*, and random kernel* R *is given above. We assume that*

$$\mathbb{E}[\log \rho_0] < 0, \tag{6.6}$$

and there exists $k > 2$, such that

$$\mathbb{E}\left[\rho_0^k\right] = 1, \text{ and } \mathbb{E}\left[\rho_0^k \log^+ \rho_0\right] < \infty. \tag{6.7}$$

Let the starting configuration be U_0. If Z_n denotes the color of the $(n+1)$-th selected ball, then there exists a non-random $\sigma^2 > 0$, such that, as $n \to \infty$,

$$\frac{Z_n - v\log n}{v^{3/2}\sigma\sqrt{\log n}} \Rightarrow \Phi. \tag{6.8}$$

PROOF. Let $(X_n)_{n\geq 0}$ be associated stochastic process, which is a RWRE on $\mathbb{Z}$ with R as the random transition kernel. From [31] we know that there exists $\sigma_1 > 0$, such that as $n \to \infty$,

$$\mathbf{P}\left(\frac{X_n - nv}{v^{3/2}\sigma_1\sqrt{n}} \leq x\right) \longrightarrow \Phi(x), \quad \text{for all } x \in \mathbb{R}. \tag{6.9}$$

Now observe that,

$$\frac{X_{\tau_n} - v\log n}{v^{3/2}\sigma_1\sqrt{\log n}} = \frac{X_{\tau_n} - v\tau_n}{v^{3/2}\sigma_1\sqrt{\tau_n}}\sqrt{\frac{\tau_n}{\log n}} + \frac{1}{\sqrt{v}\sigma_1}\frac{\tau_n - \log n}{\sqrt{\log n}}. \tag{6.10}$$

By Theorem 6.2 we know that equation (6.5) holds where $(X_n)_{n\geq 0}$ and $(\tau_n)_{n\geq 0}$ are independent. Therefore, from (2.2), and (6.9), it follows that as $n \to \infty$,

$$\mathbf{P}\left(\frac{X_{\tau_n} - v\tau_n}{v^{3/2}\sigma_1\sqrt{\log n}} \leq x\right) \longrightarrow \Phi(x), \text{ for all } x \in \mathbb{R}. \tag{6.11}$$

Define $\sigma^2 := \left(1 + \frac{1}{v\sigma_1^2}\right)$. Then (6.8) follows from (2.3) and (6.11). □

REMARK 6.4. As the above proof indicates, the conclusion in (6.8) also holds if we assume (6.6) and a slightly restrictive condition that $\mathbb{E}\left[\rho_0^k\right] < 1$, for some $k > 2$. The key idea is to show (6.9) should hold, which is essential for the computations as done in (6.10). Theorem 5 on page 10 of [9] shows that (6.9) holds.

6.2.2. *Urn Associated with a Recurrent RWRE on $\mathbb{Z}$.*

THEOREM 6.5. *Consider an infinite color random replacement scheme urn model with colors indexed by $S = \mathbb{Z}$, and random kernel R is as given above. We assume that*

$$\mathbb{E}\left[\log \rho_0\right] = 0, \tag{6.12}$$

and

$$\mathbb{P}\left(\rho_0 = 1\right) < 1. \tag{6.13}$$

Let

$$0 < \sigma^2 := \mathbb{E}\left[(\log \rho)_0^2\right] < \infty.$$

Let the starting configuration be U_0. If Z_n denotes the color of the $(n+1)$-th selected ball, then as $n \to \infty$,

$$\frac{\sigma^2 Z_n}{(\log\log n)^2} \Rightarrow G, \tag{6.14}$$

where G is a continuous distribution on $\mathbb{R}$ with probability density function g, given by

$$g(x) = \frac{2}{\pi}\sum_{k=0}^{\infty}\frac{(-1)^k}{2k+1}\exp\left(-\frac{(2k+1)^2\pi^2}{8}|x|\right), \quad x \in \mathbb{R}. \tag{6.15}$$

PROOF. It is known from [43] that as $n \to \infty$,

$$\mathbf{P}\left(\frac{\sigma^2 X_n}{(\log n)^2} \leq x\right) \longrightarrow \int_{-\infty}^{x} g(t)\, dt, \text{ for all } x \in \mathbb{R}, \tag{6.16}$$

Observe that,

$$\frac{\sigma^2 X_{\tau_n}}{(\log\log n)^2} = \frac{\sigma^2 X_{\tau_n}}{(\log \tau_n)^2}\left(\frac{\log \tau_n}{\log\log n}\right)^2. \tag{6.17}$$

By Theorem 6.2 we know that equation (6.5) holds where $(X_n)_{n\geq 0}$ and $(\tau_n)_{n\geq 0}$ are independent. Therefore, as $n \to \infty$,

$$\mathbf{P}\left(\frac{\sigma^2 X_{\tau_n}}{(\log \tau_n)^2} \leq x\right) \longrightarrow G(x), \text{ for all } x \in \mathbb{R}. \tag{6.18}$$

Also, from equation (2.2) it follows that as $n \to \infty$,

$$\frac{\log \tau_n}{\log\log n} \longrightarrow 1, \quad \mathbf{P}- \text{ a.s.} \tag{6.19}$$

Finally, (6.14) follows from (6.5), (6.18) and (6.19). □

7. Conclusion. We have presented in this paper a new method for studying finite or infinite color *balanced* urn schemes through their representation in the associated Markov chain. It turns out that for any general balanced urn scheme, the sequence of observed colors is a realization of a branching Markov chain with a modified kernel on the random recursive tree. We have shown using such representation that under fairly general conditions one can derive asymptotic of various urn schemes, which otherwise may be very difficult to find. As illustrated by several examples, essentially the asymptotic may be derived if the underlying Markov chain has a proper scaling limit. We believe that this novel approach will provide a better understanding of these new type of urn schemes.

Acknowledgments. The authors are grateful to the anonymous reviewers and the Associate Editor for their very insightful remarks, which have vastly improved the quality of the exposition. The authors also like to express their sincere gratitude to Arijit Chakrabarty, Codina Cotar, Krishanu Maulik and Tatyana Turova for various discussions they had with them at various time points.

REFERENCES

[1] ATHREYA, K. B. and KARLIN, S. (1968). Embedding of urn schemes into continuous time Markov branching processes and related limit theorems. *Ann. Math. Statist.* **39** 1801–1817. MR0232455 (38 ##780)

[2] BAGCHI, A. and PAL, A. K. (1985). Asymptotic normality in the generalized Pólya-Eggenberger urn model, with an application to computer data structures. *SIAM J. Algebraic Discrete Methods* **6** 394–405. MR791169 (86j:60025)

[3] BAI, Z.-D. and HU, F. (2005). Asymptotics in randomized urn models. *Ann. Appl. Probab.* **15** 914–940. MR2114994 (2005k:62055)

[4] BANDYOPADHYAY, A., JANSON, S. and THACKER, D. (2020). Strong convergence of infinite color balanced urns under uniform ergodicity. *J. Appl. Probab.* **57** 853–865. MR4148061

[5] BANDYOPADHYAY, A. and THACKER, D. (2014). Rate of convergence and large deviation for the infinite color Pólya urn schemes. *Statist. Probab. Lett.* **92** 232–240. MR3230498

[6] BANDYOPADHYAY, A. and THACKER, D. (2017). Pólya urn schemes with infinitely many colors. *Bernoulli* **23** 3243–3267. MR3654806

[7] BILLINGSLEY, P. (1995). *Probability and measure*, third ed. *Wiley Series in Probability and Mathematical Statistics*. John Wiley & Sons Inc., New York. MR1324786 (95k:60001)

[8] BLACKWELL, D. and MACQUEEN, J. B. (1973). Ferguson distributions via Pólya urn schemes. *Ann. Statist.* **1** 353–355. MR0362614 (50 ##15054)

[9] BOGACHEV, L. V. (2006). Random Walks in random Environments. *Encyclopedia. Math. Phys.* **4** 353–371.

[10] BOSE, A., DASGUPTA, A. and MAULIK, K. (2009). Multicolor urn models with with reducible replacement matrices. *Bernoulli* **15** 279–295. MR2546808 (2010j:60023)

[11] BOSE, A., DASGUPTA, A. and MAULIK, K. (2009). Strong laws for balanced triangular urns. *J. Appl. Probab.* **46** 571–584. MR2535833 (2010j:60024)

[12] BOURBAKI, N. (1971). *Éléments de mathématique. Topologie générale. Chapitres 1 à 4*. Hermann, Paris. MR0358652

[13] CHEN, M.-R., HSIAU, S.-R. and YANG, T.-H. (2014). A new two-urn model. *J. Appl. Probab.* **51** 590–597. MR3217787

[14] CHEN, M.-R. and KUBA, M. (2013). On generalized Pólya urn models. *J. Appl. Probab.* **50** 1169–1186. MR3161380

[15] COLLEVECCHIO, A., COTAR, C. and LICALZI, M. (2013). On a preferential attachment and generalized Pólya's urn model. *Ann. Appl. Probab.* **23** 1219–1253. MR3076683

[16] CRANE, E., GEORGIOU, N., VOLKOV, S., WADE, A. R. and WATERS, R. J. (2011). The simple harmonic urn. *Ann. Probab.* **39** 2119–2177. MR2932666

[17] CURTIS, C. W. (1993). *Linear algebra*, fourth ed. *Undergraduate Texts in Mathematics*. Springer-Verlag, New York. MR1284922

[18] DASGUPTA, A. and MAULIK, K. (2011). Strong laws for urn models with balanced replacement matrices. *Electron. J. Probab.* **16** no. 63, 1723–1749. MR2835252

[19] DEVROYE, L. (1988). Applications of the theory of records in the study of random trees. *Acta Inform.* **26** 123–130. MR969872 (89m:60027)

[20] DEVROYE, L., FAWZI, O. and FRAIMAN, N. (2012). Depth properties of scaled attachment random recursive trees. *Random Structures Algorithms* **41** 66–98. MR2943427

[21] DURRETT, R. (2010). *Probability: theory and examples*, fourth ed. *Cambridge Series in Statistical and Probabilistic Mathematics*. Cambridge University Press, Cambridge. MR2722836 (2011e:60001)

[22] FELLER, W. (1971). *An introduction to probability theory and its applications. Vol. II. Second edition*. John Wiley & Sons, Inc., New York-London-Sydney. MR0270403 (42 ##5292)

[23] FLAJOLET, P., DUMAS, P. and PUYHAUBERT, V. (2006). Some exactly solvable models of urn process theory. In *Fourth Colloquium on Mathematics and Computer Science Algorithms, Trees, Combinatorics and Probabilities*. *Discrete Math. Theor. Comput. Sci. Proc., AG* 59–118. Assoc. Discrete Math. Theor. Comput. Sci., Nancy. MR2509623 (2011b:60030)

[24] FREEDMAN, D. A. (1965). Bernard Friedman's urn. *Ann. Math. Statist* **36** 956–970. MR0177432 (31 ##1695)

[25] FRIEDMAN, B. (1949). A simple urn model. *Comm. Pure Appl. Math.* **2** 59–70. MR0030144 (10,720b)

[26] GALAMBOS, J. and SENETA, E. (1973). Regularly varying sequences. *Proc. Amer. Math. Soc.* **41** 110–116. MR0323963

[27] GOUET, R. (1997). Strong convergence of proportions in a multicolor Pólya urn. *J. Appl. Probab.* **34** 426–435. MR1447347 (98f:60065)

[28] JANSON, S. (2004). Functional limit theorems for multitype branching processes and generalized Pólya urns. *Stochastic Process. Appl.* **110** 177–245. MR2040966 (2005a:60134)

[29] JANSON, S. (2006). Limit theorems for triangular urn schemes. *Probab. Theory Related Fields* **134** 417–452. MR2226887 (2007b:60051)

[30] JANSON, S. (2018). A.s. convergence for infinite colour Pólya urns associated with random walks. (https://arxiv.org/abs/1803.04207).

[31] KESTEN, H., KOZLOV, M. V. and SPITZER, F. (1975). A limit law for random walk in a random environment. *Compositio Math.* **30** 145–168. MR0380998 (52 ##1895)

[32] KUBA, M. and WAGNER, S. (2010). On the distribution of depths in increasing trees. *Electron. J. Combin.* **17** Research Paper 137, 9. MR2729386

[33] LARUELLE, S. and PAGÈS, G. (2013). Randomized urn models revisited using stochastic approximation. *Ann. Appl. Probab.* **23** 1409–1436. MR3098437

[34] MAHMOUD, H. M. (2009). *Pólya urn models*. *Texts in Statistical Science Series*. CRC Press, Boca Raton, FL. MR2435823

[35] MAILLER, C. and MARCKERT, J.-F. (2017). Measure-valued Pólya urn processes. *Electron. J. Probab.* **22** Paper No. 26, 33. MR3629870

[36] MEYN, S. and TWEEDIE, R. L. (2009). *Markov chains and stochastic stability*, Second ed. Cambridge University Press, Cambridge. MR2509253

[37] MOON, J. W. (1974). The distance between nodes in recursive trees. In *Combinatorics (Proc. British Combinatorial Conf., Univ. Coll. Wales, Aberystwyth, 1973)* 125–132. London Math. Soc. Lecture Note Ser., No. 13. MR0357186

[38] PEMANTLE, R. (1990). A time-dependent version of Pólya's urn. *J. Theoret. Probab.* **3** 627–637. MR1067672 (91i:60030)

[39] PEMANTLE, R. (2007). A survey of random processes with reinforcement. *Probab. Surv.* **4** 1–79. MR2282181 (2007k:60230)

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

[41] RUDIN, W. (1976). *Principles of mathematical analysis*, Third ed. McGraw-Hill Book Co., New York-Auckland-Düsseldorf. MR0385023

[42] SENETA, E. (2006). *Non-negative matrices and Markov chains*. *Springer Series in Statistics*. Springer, New York. MR2209438

[43] SINAĬ, Y. G. (1982). The limit behavior of a one-dimensional random walk in a random environment. *Teor. Veroyatnost. i Primenen.* **27** 247–258. MR657919 (83k:60078)

[44] SMYTHE, R. T. and MAHMOUD, H. M. (1994). A survey of recursive trees. *Teor. Ĭmovīr. Mat. Stat.* **51** 1–29. MR1445048

[45] ZEITOUNI, O. (2004). Random walks in random environment. In *Lectures on probability theory and statistics*. *Lecture Notes in Math.* **1837** 189–312. Springer, Berlin. MR2071631 (2006a:60201)