PÓLYA URNS AND OTHER REINFORCEMENT PROCESSES

CÉCILE MAILLER

What is a Pólya urn? A d-colour Pólya urn (for $d \geqslant 2$) is a Markov process $(U(n))_{n\geqslant 0}$ whose distribution depends on two parameters:

- the initial composition $U(0) \in \mathbb{N}^d$,
- the replacement matrix $R = (R_{ij})_{1\leqslant i,j\leqslant d}$, a d-dimensional matrix whose coefficients are integer-valued.

The distribution of $(U(n))_{n\geqslant 0}$ is defined recursively as follows: For all $n \geqslant 0$, let $\xi(n+1)$ be a random variable in $\{1, \dots, d\}$ with distribution

$$\mathbb{P}(\xi(n+1) = i|U(n)) = \frac{U_i(n)}{\|U(n)\|_1}, \quad (\forall 1 \leqslant i \leqslant d),$$

where$\|U(n)\|_1 = \sum_{i=1}^d U_i(n)$. Set $U(n+1) = U(n) + R_{\xi(n+1)}$, where, for all $1 \leqslant i \leqslant d$, R_i is the i-th line of the matrix R.

The idea is that $(U(n))_{n\geqslant 0}$ represents the evolution of the contents of an urn that contains balls of possible colours $1, 2, \dots, d$. For all $n \geqslant 0$ and $1 \leqslant i \leqslant d$, $U_i(n)$ is the number of balls of colour i in the urn at time n and, at each time step n, we pick a ball uniformly at random in the urn, let $\xi(n)$ denote its colour, and replace it in the urn together with an additional $R_{\xi(n)j}$ balls of colour j, for all $1 \leqslant j \leqslant d$.

A few comments on the initial composition and replacement matrix. We have asked that the coefficients of $U(0)$ and R are integers. This is not necessary for the mathematical definition of the model (and for most of the results), although the interpretation in terms of "balls in an urn" breaks down if these are not integers. That said, one can think of coloured dust in an urn instead.

The coefficients of R can be negative, meaning that we remove balls from the urn instead of adding balls to the urn. This is fine as long as the number of balls of each colour stays positive at all times. If the number of balls of one colour becomes negative, we say that the urn becomes extinct and the process stops. One can come up with conditions on $U(0)$ and R that ensure that the urn almost surely never comes extinct (see, e.g., [Mah08]): for example, one can ask that

$$\begin{cases} R_{ii} \geqslant -1 & \text{for all } 1 \leqslant i \leqslant d, \text{ and} \\ R_{ij} \geqslant 0 & \text{for all } 1 \leqslant i \neq j \leqslant d. \end{cases} \tag{0.1}$$

Indeed, $R_{ii} = -1$ means that we remove from the urn the ball that we picked at random. An urn that almost surely never becomes extinct is called "tenable". One sufficient condition for tenability is that, for all $1 \leqslant i \leqslant d$, there exists κ_i such that $R_{ii} = -\kappa_i$ and, for all $1 \leqslant j \leqslant d$, $R_{ij} \geqslant 0$ is divisible by κ_i. For simplicity, in this course, we assume that this sufficient condition always holds.

What questions are we interested in? In this course, we ask what is the composition of the urn as n goes to infinity, i.e. we aim at proving limiting theorems for $U(n)$ as $n \uparrow \infty$. Does $U(n)/n$ converge as $n \uparrow \infty$? Does it converge in probability? almost surely? If it converges, what is its limit? And what are the fluctuations around this limit? Naturally, the answers to these questions depend on the parameters $U(0)$ and R of the Pólya urn.

In this course, we will consider the following two cases:

(1) The replacement matrix is $R = S\mathrm{Id}$. We call this case "the identity case".
(2) The replacement matrix is irreducible. We call this case "the irreducible case".

We will see that these two cases lead to radically different behaviours: In the identity case, $U(n)/n$ converges almost surely to a random limit whose distribution depends on the initial composition of the urn. In the irreducible case, $U(n)/n$ converges almost surely to a limit that does not depend on the initial composition of the urn. We will show these "law of large numbers" results and also look at the fluctuations around these almost sure limits.

In the second part of the course, we will look at extending Pólya urns to the case when the set of colours is infinite.

1. Finitely-many colour Pólya urns

1.1. **The identity case.** In this section, we consider the case when $R = S\mathrm{Id}$. Without loss of generality, we assume that, for all $1 \leqslant i \leqslant d$, $\alpha_i := U_i(0) \geqslant 1$. We show convergence of the composition of the urn when time goes to infinity, as well as convergence of the fluctuations of the composition around its almost sure limit.

1.1.1. *A law of large numbers.* The following theorem dates back to Markov [Mar17] for $S = 1$, Eggenberger and Pólya [EP23] also for $S = 1$.

Theorem 1.1. *Assume that $R = S\mathrm{Id}$ and, for all $1 \leqslant i \leqslant d$, $\alpha_i := U_i(0) \geqslant 1$. Then, almost surely as $n \to +\infty$,*

$$\frac{U(n)}{Sn} \to V = (V_1, \dots, V_d),$$

where V is Dirichlet-distributed with parameter $({}^{\alpha_1}/_S, \dots, {}^{\alpha_d}/_S)$.

We recall that the density of the Dirichlet distribution of parameter $(\nu_1, \dots, \nu_d)$ is given by

$$\frac{\Gamma(\nu_1 + \dots + \nu_d)}{\Gamma(\nu_1) \dots \Gamma(\nu_d)} \prod_{i=1}^{d} x_i^{\nu_i - 1} \mathrm{d}\Sigma(x_1, \dots, x_d),$$

where $\mathrm{d}\Sigma(x_1, \dots, x_d)$ is the Lebesgue measure on the simplex

$$\Sigma = \Big\{(x_1, \dots, x_d) \in [0,1]^d \colon \sum_{i=1}^{d} x_i = 1\Big\}.$$

In particular, the Dirichlet distribution of parameter $(1, \dots, 1)$ is the uniform distribution on Σ. Two well-known particular cases of Theorem 1.1 are the following:

Corollary 1.2. *Assume that $d = 2$ and $R = \mathrm{Id}$.*

(i) *If $\alpha_1 = \alpha_2 = 1$, then*

$$\frac{U(n)}{n} \to (X, 1 - X),$$

almost surely as $n \uparrow \infty$, where $X \sim \mathrm{Unif}(0,1)$.

(ii) *If $\alpha_1, \alpha_2 \geqslant 1$, then*

$$\frac{U(n)}{n} \to (B, 1 - B),$$

almost surely as $n \uparrow \infty$, where B is a Beta random variable of parameter (α_1, α_2).

Proof of Theorem 1.1. For all $n \geqslant 0$,

$$\mathbb{E}[U(n+1)|U(n)] = \mathbb{E}[U(n) + Se_{\xi(n+1)}],$$

where $(e_1, \dots, e_d)$ is the canonical basis of $\mathbb{R}^d$ (i.e., for all $1 \leqslant i \leqslant d$, all coordinates of e_i are null except the i-th one, which equals 1), and $\xi(n+1)$ is the colour of the ball drawn at time $n+1$. We thus get

$$\mathbb{E}[U(n+1)|U(n)] = U(n) + S\mathbb{E}[e_{\xi(n+1)}] = U(n) + S\sum_{i=1}^{d} \frac{U_i(n)}{\|U(n)\|_1}\, e_i = \Big(1 + \frac{S}{\|U(n)\|_1}\Big)U(n).$$

Thus, $(M_n = U(n)/\|U(n)\|_1)_{n\geqslant 0}$ is a martingale for its natural filtration. Because it is non-negative, by the martingale convergence theorem, it converges almost surely to an almost finite random variable V. By definition, $\|U(n)\|_1 = \|U(0)\|_1 + nS$ for all $n \geqslant 0$. Thus, almost surely as $n \uparrow \infty$,

$$\frac{U(n)}{Sn} = \frac{U(n)}{\|U(n)\|_1} \cdot \frac{\|U(n)\|_1}{Sn} \to V.$$

To prove that V is Dirichlet distributed with parameter $({}^{\alpha_1}\!/\!{}_S, \dots, {}^{\alpha_d}\!/\!{}_S)$, one can calculate moments of M_n and show that they converge to the moments of the Dirichlet distribution. This is done in [CMP15, Section 6].

In this course, we only prove that the distribution of X in Corollary 1.2(i) is indeed uniform on $[0,1]$. In fact, we prove that, for all $n \geqslant 1$, $1 \leqslant k \leqslant n+1$,

$$\mathbb{P}(U_1(n) = k) = \frac{1}{n+1}. \tag{1.1}$$

In other words, the distribution of $U_1(n)$ is uniform on $\{1, \dots, n+1\}$ for all $n \geqslant 1$, which indeed implies that X is uniform on $[0,1]$, as claimed. One can easily prove (1.1) by induction. Instead, we discuss here how one could have guessed this formula from scratch: to have k balls of colour 1 in the urn at time n, one needs to have picked $(k-1)$ times a ball of colour 1 and $(n-k+1)$ times a ball of colour 2. The probability that we draw a ball of colour 1 at times $1, \dots, k-1$ and a ball of colour 2 at times $k, \dots, n-1$ is given by

$$\frac{1}{2}\cdot\frac{2}{3}\cdots\frac{k-1}{k+1}\cdot\frac{1}{k+2}\cdot\frac{2}{k+3}\cdots\frac{n-k+1}{n+1} = \frac{(k-1)!(n-k)!}{(n+1)!}.$$

Now note that, if the draws happened in a different order, "only the order of the numerators change", which means that the probability above is the same, independently of the order in which the $k-1$ balls of colour 1 and $n-k$ balls of colour 2 have been drawn. Since there are $\binom{n}{k-1}$ such orders, we get

$$\mathbb{P}(U_1(n) = k) = \binom{n}{k-1}\frac{(k-1)!(n-k+1)!}{(n+1)!} = \frac{1}{n+1},$$

as claimed in (1.1). This concludes the proof in the case of Corollary 1.2(i). The proof in the case of Corollary 1.2(ii) is similar, only more technical. □

The following result is useful when trying to simulate a Pólya urn easily in the identity case. In the case of Corollary 1.2(i), it is an exercise in the book of Williams [Wil91]:

Theorem 1.3. *Fix $\alpha_1, \dots, \alpha_d \geqslant 1$ and $S \geqslant 1$. Define the d-dimensional process $(Z(n))_{n\geqslant 0}$ as follows: let V be a Dirichlet random variable of parameter $({}^{\alpha_1}\!/\!{}_S, \dots, {}^{\alpha_d}\!/\!{}_S)$ and let $(X_m)_{m\geqslant 1}$ be a sequence of random variables such that, for all $k \geqslant 1$, for all $1 \leqslant i \leqslant d$,*

$$\mathbb{P}(X_k = e_i|V) = V_i.$$

Let $Z_i(0) = \alpha_i \geqslant 1$ *for all* $1 \leqslant i \leqslant d$ *and, for all* $n \geqslant 1$,

$$Z(n) = Z(0) + S \sum_{k=1}^{n} X_k.$$

Then, $(Z(n))_{n\geqslant 0}$ *is a Pólya urn of initial composition* $(\alpha_1, \dots, \alpha_d)$ *and replacement matrix* $S\mathrm{Id}$.

NB: By the law of large numbers, it is clear that, conditionally on V, $Z(n)/(nS) \to V$ almost surely as $n \uparrow \infty$, i.e.

$$\mathbb{P}\Big(\frac{Z(n)}{nS} \to V \,\Big|\, V\Big) = 1.$$

By the tower rule, this implies that $Z(n)/(nS) \to V$ almost surely as $n \uparrow \infty$, which confirms the law of large numbers stated in Theorem 1.1.

Proof. Let $(U(n))_{n\geqslant 0}$ be a Pólya urn of initial composition $(\alpha_1, \dots, \alpha_d)$ and replacement matrix $S\mathrm{Id}$. We aim to prove that, for all $n \geqslant 0$, $k_1, \dots, k_d \geqslant 0$,

$$\mathbb{P}\big(Z_i(n) = \alpha_i + Sk_i\,(\forall 1 \leqslant i \leqslant d)\big) = \mathbb{P}\big(U_i(n) = \alpha_i + Sk_i\,(\forall 1 \leqslant i \leqslant d)\big). \tag{1.2}$$

This indeed implies that $(Z(n))_{n\geqslant 0}$ is distributed as $(U(n))_{n\geqslant 0}$, as claimed.

To prove (1.2), we give explicit formulae for the LHS and RHS and show that they are indeed equal. For the RHS, we use a similar argument as in the proof of (1.1): for all $n \geqslant 0$, $k_1, \dots, k_d \geqslant 0$,

$$\mathbb{P}(U_i(n) = \alpha_i + Sk_i) = \frac{n!}{k_1! \dots k_d!} \frac{\prod_{i=1}^d \alpha_i(\alpha_i + S) \cdots (\alpha_i + (k_i - 1)S)}{\prod_{i=0}^{n-1} (\bar\alpha + iS)} \mathbf{1}_{k_1+\dots+k_d=n},$$

where we have set $\bar\alpha = \sum_{i=1}^d \alpha_i$. Dividing both numerator and denominator by S gives

$$\begin{aligned}\mathbb{P}(U_i(n) = U_i(0) + Sk_i\,(\forall 1 \leqslant i \leqslant d)) &= \frac{n!}{k_1! \dots k_d!} \frac{\prod_{i=1}^d (\alpha_i/S)(\alpha_i/S + 1) \cdots (\alpha_i/S + k_i - 1)}{\prod_{i=0}^{n-1} (\bar\alpha/S + i)} \mathbf{1}_{k_1+\dots+k_d=n} \\ &= \frac{n!}{k_1! \dots k_d!} \cdot \frac{\Gamma(\bar\alpha/S)}{\Gamma(\bar\alpha/S + n)} \cdot \frac{\prod_{i=1}^d \Gamma(\alpha_i/S + k_i)}{\prod_{i=1}^d \Gamma(\alpha_i/S)} \mathbf{1}_{k_1+\dots+k_d=n}. \qquad (1.3)\end{aligned}$$

Now note that, for all $n \geqslant 0$, $k_1, \dots, k_d \geqslant 0$,

$$\begin{aligned}\mathbb{P}\big(Z_i(n) = \alpha_i + Sk_i\,(\forall 1 \leqslant i \leqslant d)\big) &= \mathbb{E}\big[\mathbb{P}\big(Z_i(n) = \alpha_i + Sk_i\,(\forall 1 \leqslant i \leqslant d)\big|V\big)\big] \\ &= \mathbb{E}\Big[\frac{n!}{k_1! \dots k_d!} \prod_{i=1}^d V_i^{k_i} \mathbf{1}_{k_1+\dots+k_d=n}\Big] \\ &= \frac{n!}{k_1! \dots k_d!} \mathbb{E}\Big[\prod_{i=1}^d V_i^{k_i}\Big] \mathbf{1}_{k_1+\dots+k_d=n}.\end{aligned}$$

It is known that, if V is Dirichlet-distributed with parameter $(\nu_1, \dots, \nu_d)$, then

$$\mathbb{E}\Big[\prod_{i=1}^d V_i^{k_i}\Big] = \frac{\Gamma(\bar\nu)}{\Gamma(\bar\nu + |k|)} \prod_{i=1}^d \frac{\Gamma(\nu_i + k_i)}{\Gamma(\nu_i)},$$

where $\bar\nu = \sum_{i=1}^d \nu_i$ and $|k| = \sum_{i=1}^d k_i$. We thus get

$$\mathbb{P}\big(Z_i(n) = \alpha_i + Sk_i\,(\forall 1 \leqslant i \leqslant d)\big) = \frac{n!}{k_1! \dots k_d!} \cdot \frac{\Gamma(\bar\alpha/S)}{\Gamma(\bar\alpha/S + n)} \cdot \frac{\prod_{i=1}^d \Gamma(\alpha_i/S + k_i)}{\prod_{i=1}^d \Gamma(\alpha_i/S)} \mathbf{1}_{k_1+\dots+k_d=n}.$$

Therefore, by (1.3), for all $n \geqslant 0$, for all $k_1, \dots, k_d \geqslant 1$,

$$\mathbb{P}\big(Z_i(n) = \alpha_i + Sk_i\,(\forall 1 \leqslant i \leqslant d)\big) = \mathbb{P}\big(U_i(n) = \alpha_i + Sk_i\,(\forall 1 \leqslant i \leqslant d)\big),$$

and thus, $(Z(n))_{n\geqslant 0}$ is indeed distributed as the Pólya urn of initial composition $(\alpha_1, \dots, \alpha_d)$ and replacement matrix $S\mathrm{Id}$, as claimed. □

1.1.2. *A central limit theorem.* When we have a law of large numbers as in Theorem 1.1, it is natural to try and prove a central limit theorem for the fluctuations of the random quantity around its almost sure limit. In the identity case, these fluctuations are given by a Gaussian of random variance, function of the almost sure limit V:

Theorem 1.4 (see, e.g. [Mül]). *Assume that $(U(n))_{n\geqslant 0}$ is a Pólya urn of initial composition $(\alpha_1, \dots, \alpha_d)$ and replacement matrix $R = S\mathrm{Id}$ for some $\alpha_1, \dots, \alpha_d, S \geqslant 1$. Then, in distribution as $n \uparrow \infty$,*

$$\frac{U(n) - nSV}{S\sqrt{n}} \Rightarrow \mathcal{N}(0, \Sigma^2), \tag{1.4}$$

where

$$\Sigma^2 = \begin{pmatrix} V_1(1-V_1) & -V_1V_2 & \dots & -V_1V_d \\ -V_1V_2 & V_2(1-V_2) & & -V_2V_d \\ \vdots & & \ddots & \vdots \\ -V_1V_d & -V_2V_d & \dots & V_d(1-V_d) \end{pmatrix}. \tag{1.5}$$

NB: Equation (1.4) means that, for all Borel sets $B \subset \mathbb{R}^d$,

$$\mathbb{P}\left(\frac{U(n) - nSV}{S\sqrt{n}} \in B\right) \to \mathbb{E}\left[\frac{1}{(2\pi)^{d/2}\sqrt{\det(\Sigma)}} \int_B \mathrm{e}^{-x\Sigma^2 x^T/2} \mathrm{d}x\right].$$

Proof. This is an easy consequence of Theorem 1.3: indeed, we have that, in distribution, $U(n) = Z(n)$, where $Z(n)$ is defined as in Lemma 1.3. Conditionally on V, $Z(n) = (\alpha_1, \dots, \alpha_d) + S\sum_{k=1}^d X_k$, where $\mathbb{P}(X_k = e_i) = V_i$ for all $k \geqslant 1$ and $1 \leqslant i \leqslant d$. By the central limit theorem, conditionally on V,

$$\frac{Z(n) - nSV}{S\sqrt{n}} \Rightarrow \mathcal{N}(0, \Sigma^2),$$

where $\Sigma^2 = \mathrm{Cov}(X|V)$, where X is a copy of X_1. Because $\mathbb{E}[X] = V$

$$\begin{aligned} \mathrm{Cov}(X|V) = \mathbb{E}\big[(X-V)^T(X-V)|V\big] &= \sum_{i=1}^d V_i(e_i - V)^T(e_i - V) \\ &= \sum_{i=1}^d V_i(e_ie_i^T - e_i^TV - V^Te_i + V^TV) \\ &= \sum_{i=1}^d V_ie_ie_i^T - \left(\sum_{i=1}^d V_ie_i^T\right)V - V^T\sum_{i=1}^d V_ie_i + V^TV\sum_{i=1}^d V_i \\ &= \sum_{i=1}^d V_ie_ie_i^T - V^TV - V^TV + V^TV = \sum_{i=1}^d V_ie_ie_i^T - V^TV, \end{aligned}$$

and this can indeed be written as in (1.5). □

1.2. **The irreducible case.** In this section, we look at the case when the replacement matrix R is irreducible, i.e. for all $1 \leqslant i, j \leqslant d$, there exists $n = n(i,j)$ such that $R^n_{i,j} > 0$. In other words, for all $1 \leqslant i, j \leqslant d$, if $U(0) = e_i$ (where e_i is the vector whose coordinates are all equal to 0 except the i-th which equals 1), then there is a positive probability to see a ball of colour j at some (finite) time in the urn.

First note that, because we have assumed that all non-diagonal coefficients of R are non-negative. Thus, there exists $\kappa \geqslant 0$ such that $R + \kappa\mathrm{Id}$ has non-negative coefficients. It is also irreducible (because $(R + \kappa\mathrm{Id})^n = \sum_{i=1}^n \binom{n}{i}\kappa^{n-i}R^i$ for all $n \geqslant 1$). Thus, by Perron-Frobenius's theorem, the spectral radius ρ of $R + \kappa\mathrm{Id}$ is larger than κ and a simple eigenvalue of $R + \kappa\mathrm{Id}$. Furthermore, there exists v a left-eigenvector of $R + \kappa\mathrm{Id}$ whose coefficients are all positive such that $\|v\|_1 = 1$. From this, we deduce that $\lambda = \rho - \kappa > 0$ is the eigenvalue of R with largest real part and $vR = \lambda v$. We call λ and v the dominant eigenvalue and dominant left-eigenvector of R, respectively.

1.2.1. *A law of large numbers.* The following "law of large numbers" is due to Athreya and Karlin [AK68]:

Theorem 1.5. *Fix $d \geqslant 2$. Let $(U(n))_{n\geqslant 0}$ be the d-colour Pólya urn of initial composition $(\alpha_1, \dots, \alpha_d)$ and replacement matrix R. Assume that $\bar\alpha := \sum_{i=1}^d \alpha_i \geqslant 1$ and R is irreducible. Let $\lambda > 0$ and v be, respectively, the dominant eigenvalue and dominant left-eigenvector of R. Then, almost surely as $n \uparrow \infty$,*

$$\frac{U(n)}{n} \to \lambda v.$$

Before proving this result, we first make a few comments and in particular compare this behaviour of "irreducible" urns with the behaviour of the identity urns of Section 1.1. We recall that, in the identity case, $U(n)$ also satisfies a law of large numbers: see Theorem 1.1.

The similarity between Theorems 1.1 and 1.5 is that, in both cases, $U(n)/n$ converges almost surely as $n \uparrow \infty$, which is why we call both these results "law of large numbers". However, these two results are in fact drastically different. Indeed:

- The limit of $U(n)/n$ is deterministic (λv) in the irreducible case and random in the identity case (a random Dirichlet-distributed vector V).
- This limit does not depend on the initial composition in the irreducible case, while the distribution of V depends on the initial distribution in the identity case.

NB: Note that Theorem 1.5 implies that $\|U(n)\|_1/n \to \lambda$ almost surely as $n \uparrow \infty$ because $\|v\|_1 = 1$, by assumption. In other words, the total number of balls in the urn at time n grows as λn as $n \uparrow \infty$.

To prove Theorem 1.5, we embed the process into continuous time: the embedding is a multi-type Galton-Watson process and martingale theory allows us to study this process precisely. Embedding into continuous time has the advantage to give more independence, but the price to pay is that some work needs to be done to translate the results back in discrete time.

1.2.2. *Embedding of an urn into continuous time.* Given a replacement matrix R and an initial composition $\alpha = (\alpha_1, \dots, \alpha_d)$, we define the continuous-time, multi-type branching process $(X(t))_{t\geqslant 0}$ as follows: $X(0) = \alpha$, meaning that, at time 0, there are α_i particles of type i alive in the system. Each particle reproduces (or "splits") independently from the rest at rate 1, and at a reproduction event triggered by a particle of type i, we add to the system $R_{i,j}$ particles of type j, for all $1 \leqslant i, j \leqslant d$. We call X the continuous-time urn process of initial composition α and replacement matrix R.

Proposition 1.6. *Let $\alpha = (\alpha_1, \dots, \alpha_d)$ and R be a $d \times d$ replacement matrix. Let $(U(n))_{n\geqslant 0}$ be the Pólya urn of replacement matrix R and initial composition α. Let $(X(t))_{t\geqslant 0}$ be the continuous-time* *urn process of replacement matrix R and initial composition α. Let $\tau_0 = 0$ and, for all $n \geqslant 1$, τ_n be the time when the n-th split in the process $(X(t))_{t\geqslant 0}$. Then, in distribution,*

$$(U(n))_{n\geqslant 0} = (X(\tau_n))_{n\geqslant 0}.$$

1.2.3. *A law of large numbers for the continuous-time urn process.* To prove Theorem 1.5, we first prove the following law of large numbers for the continuous-time process X defined in Section 1.2.2:

Theorem 1.7. *Let X be the continuous-time urn process of initial composition $\alpha = (\alpha_1, \ldots, \alpha_d)$ and replacement matrix R. Under the assumptions of Theorem 1.5 on α and R, and using the same notation for λ and v, almost surely as $t \uparrow \infty$,*

$$\mathrm{e}^{-\lambda t} X(t) \to W v,$$

where W is an almost surely finite random variable such that $\mathbb{P}(W = 0) < 1$.

For the proof, we start with the following lemma:

Lemma 1.8. *The continuous-time process $(X(t)\mathrm{e}^{-tR})_{t\geqslant 0}$ is a (vector-valued) martingale.*

Proof. We first show that, for all $t \geqslant 0$,

$$\mathbb{E}[X(t)] = X(0)\mathrm{e}^{tR}. \tag{1.6}$$

For all $1 \leqslant i \leqslant d$, we let $X^{(i)}$ be the urn process of initial composition e_i and replacement matrix R. We start by calculating $\mathbb{E}[X^{(i)}(t)]$, for all $t \geqslant 0$. We look at the time when the ball in the urn at time zero splits (with probability e^{-t}, the initial ball hasn't split yet at time t): for all $t \geqslant 0$,

$$\mathbb{E}[X^{(i)}(t)] = e_i \mathrm{e}^{-t} + \int_0^t \mathrm{e}^{-s}\mathrm{d}s \mathbb{E}\bigg[\sum_{j=1}^{d} \sum_{k=1}^{R_{i,j}+\delta_{i,j}} X^{(j,k)}(t-s)\bigg],$$

where, for all $1 \leqslant j \leqslant d$, $(X^{(j,k)})_{k\geqslant 1}$ is a sequence of i.i.d. copies of $X^{(j)}$, and the double-indexed sequence $(X^{(j,k)})_{i,j\geqslant 1}$ is a sequence of independent processes. This gives

$$\mathbb{E}\big[X^{(i)}(t)\big] = e_i \mathrm{e}^{-t} + \int_0^t \sum_{j=1}^{d} (R_{i,j}+\delta_{i,j})\mathbb{E}\big[X^{(j)}(t-s)\big]\mathrm{e}^{-s}\mathrm{d}s = e_i \mathrm{e}^{-t} + \mathrm{e}^{-t}\int_0^t \sum_{j=1}^{d} (R_{i,j}+\delta_{i,j})\mathbb{E}\big[X^{(j)}(s)\big]\mathrm{e}^{s}\mathrm{d}s.$$

So, if we let $u_i(t) := \mathbb{E}\big[X^{(i)}(t)\big]$ (a horizontal d-dimensional vector) and $u(t)$ be the $d \times d$ matrix whose lines are $u_1(t), \ldots, u_d(t)$, we get

$$u(t)\mathrm{e}^t = \mathrm{Id} + \int_0^t u(s)\mathrm{e}^s(\mathrm{Id} + A)\mathrm{d}s,$$

where we have set $A = R^T$. Differentiating in t, we get $(u'(t) + u(t))\mathrm{e}^t = u(t)\mathrm{e}^t(\mathrm{Id} + A)$, i.e. $u'(t) = u(t)A$ for all $t \geqslant 0$, which implies $u(t) = u(0)\mathrm{e}^{tA} = \mathrm{e}^{tA}$, because $u(0) = \mathrm{Id}$. Now, for a general urn process X of replacement matrix R, for all $t \geqslant 0$,

$$\mathbb{E}[X(t)] = \sum_{i=1}^{d} X_i(0)u_i(t) = X(0)u(t)^T = X(0)\mathrm{e}^{tR},$$

which concludes the proof of (1.6).

For all $t \geqslant 0$, we let $\mathcal{F}_t$ be the filtration generated by $(X(s))_{s\in[0,t]}$. By definition of the continuous-time process X, for all $s, t \geqslant 0$,

$$X(s+t) = \sum_{i=1}^{d} \sum_{j=1}^{X_i(s)} X^{(i,j)}(t).$$

Thus, for all $s, t \geqslant 0$,

$$\mathbb{E}[X(s+t)|\mathcal{F}_s] = \mathbb{E}\bigg[\sum_{i=1}^{d}\sum_{j=1}^{X_i(s)} X^{(i,j)}(t)\bigg|\mathcal{F}_s\bigg] = \sum_{i=1}^{d}\sum_{j=1}^{X_i(s)} \mathbb{E}\big[X^{(i,j)}(t)\big]$$
$$= \sum_{i=1}^{d} X_i(s)\mathbb{E}\big[X^{(i)}(t)\big] = \sum_{i=1}^{d} X_i(s)e_i\mathrm{e}^{tR} = X(s)\mathrm{e}^{tR}.$$

This implies that, for all $s, t \geqslant 0$,

$$\mathbb{E}[X(s+t)\mathrm{e}^{-(s+t)R}|\mathcal{F}_s] = X(s)\mathrm{e}^{-sR},$$

and concludes the proof that $(X(t)\mathrm{e}^{-tR})_{t\geqslant 0}$ is a martingale. □

To prove Theorem 1.7, we follow the proof of Janson [Jan04, Section 9]; however, to keep things simple, we assume that R is diagonalisable. In the general case, one needs to use the Jordan decomposition of R, and treat projections on each (generalised) eigenspace separately; eigenspaces of dimensions at least two are a bit trickier to handle, which is why we only look at eigenspaces of dimension 1 here. Under the assumption that R is diagonalisable, in addition to $v = v_1$, R admits $d-1$ left-eigenvectors $v_2, \dots, v_d$ associated to $d-1$ eigenvalues $\lambda_2, \dots, \lambda_d$. Because λ is the dominating eigenvalue of R, $\mathrm{Re}(\lambda_i) < \lambda$ for all $2 \leqslant i \leqslant d$. Given $x \in \mathbb{R}^d$, we let $\pi_i(x)$ be the i-th coordinate of x in the basis $(v_1, \dots, v_d)$; in other words, $x = \sum_{i=1}^d \pi_i(x)v_i$.

We will now prove that:

- If $\mathrm{Re}(\lambda_i) > \lambda/2$, then $\mathrm{e}^{-\lambda_i t}\pi_i(X(t))$ converges almost surely and in L^2 to an almost surely finite random variable. (See Lemma 1.9.)
- If $\mathrm{Re}(\lambda_i) \leqslant \lambda/2$, then $\mathrm{e}^{-\lambda t/2}\pi_i(X(t))$ converges in distribution to a Gaussian random variable (with an additional polynomial normalisation factor if $\mathrm{Re}(\lambda_i) = \lambda/2$).

We start with the "large" eigenvalues:

Lemma 1.9. *If $1 \leqslant i \leqslant d$ is such that $\mathrm{Re}(\lambda_i) > \lambda/2$, then,*

$$\mathrm{e}^{-\lambda_i t}\pi_i(X(t)) \to W_i \quad \textit{almost surely and in } L^2.$$

Furthermore, $\mathrm{e}^{-\lambda_i t}\pi_i(X(t)) = \mathbb{E}[W_i|\mathcal{F}_t]$, almost surely for all $t \geqslant 0$.

Proof. By Doob's L^2 martingale convergence theorem, it is enough to prove that $\big(\mathrm{e}^{-\lambda_i t}\pi_i(X(t))\big)_{t\geqslant 0}$ is bounded in L^2. To do so, we let

$$Z_k(t) = \sum_{\tau_{k,\ell}\leqslant t} \|\mathrm{e}^{-\tau_{k,\ell}R}\pi_\ell(R_k)\|_2^2,$$

and $\tau_{k,\ell}$ be the time when the ℓ-th split of a ball of colour k occurs, for all $1 \leqslant k \leqslant d$ and $\ell \geqslant 1$. With this notation, the quadratic variation of $(\pi_i(X(t))\mathrm{e}^{-tR})_{t\geqslant 0}$ is given by

$$[\pi_i(X(t))\mathrm{e}^{-tR}, \pi_i(X(t))\mathrm{e}^{-tR}]_t = \sum_{k=1}^{d} Z_k(t) \quad (\forall t \geqslant 0),$$

and, for all $t \geqslant 0$,

$$\mathbb{E}\big[\|\pi_i(X(t))\mathrm{e}^{-tR}\|_2^2\big] = \mathbb{E}\big[[\pi_i(X(t))\mathrm{e}^{-tR}, \pi_i(X(t))\mathrm{e}^{-tR}]_t\big] + \mathbb{E}\big[\|\pi_i(X(0))\|_2^2\big]$$
$$= \sum_{k=1}^{d} Z_k(t) + \mathbb{E}\big[|\pi_i(X(0))|^2\big]. \tag{1.7}$$

Because the second summand is constant, we focus on the first summand: For all $1 \leqslant k \leqslant d$,

$$\left(Z_k(t) - \int_0^t \|\mathrm{e}^{-sR}\pi_i(R_k)\|_2^2 X_k(s)\mathrm{d}s\right)_{t\geqslant 0}$$

is a martingale of mean zero, which implies, in particular,

$$\mathbb{E}[Z_k(t)] = \int_0^t \|\mathrm{e}^{-sR}\pi_i(R_k)\|_2^2 \mathbb{E}[X_k(s)]\mathrm{d}s.$$

Now note that $\mathrm{e}^{-sR}\pi_i(R_i) = \mathrm{e}^{-s\lambda_i}\pi_i(R_k)$, which implies

$$\mathbb{E}[Z_k(t)] = \int_0^t \mathrm{e}^{-2s\lambda_i}\|\pi_i(R_k)\|_2^2 \mathbb{E}[X_k(s)]\mathrm{d}s.$$

By (1.6), $\mathbb{E}[X_k(s)] = (X(0)\mathrm{e}^{Rs})_k = \mathcal{O}(\mathrm{e}^{\lambda s})$. Thus, the integrand satisfies

$$\mathrm{e}^{-2s\lambda_i}\|\pi_i(R_k)\|_2^2 \mathbb{E}[X_k(s)] = \mathcal{O}(\mathrm{e}^{(\lambda - 2\mathrm{Re}(\lambda_i))s}),$$

and is integrable because we have assumed $2\mathrm{Re}(\lambda_i) > \lambda$. This implies that $\sup_{t\geqslant 0}\mathbb{E}[Z_k(t)] < \infty$, and thus, by (1.7),

$$\sup_{t\geqslant 0}\mathbb{E}\big[\|\pi_i(X(t))\mathrm{e}^{-tR}\|_2^2\big] < \infty,$$

as desired. □

We now prove that the limits in Lemma 1.9 are not almost surely zero:

Lemma 1.10. *For all $1 \leqslant i \leqslant d$ such that $2\mathrm{Re}(\lambda_i) > \lambda$, $\mathbb{P}(W_i = 0) < 1$.*

Proof. We reason by contradiction and assume that $W_i = 0$ almost surely. Because, by Lemma 1.9, $\mathbb{E}[W_i|\mathcal{F}_t] = \pi_i(X(t)\mathrm{e}^{-tR})$ for all $t \geqslant 0$, we get that, for all $t \geqslant 0$,

$$\pi_i\big(X(t)\mathrm{e}^{-tR}\big) = 0 \quad \Rightarrow \quad \pi_i(X(t)) = 0,$$

where the implication follows from the fact that $\pi_i\big(X(t)\mathrm{e}^{-tR}\big) = \mathrm{e}^{-t\lambda_i}\pi_i\big(X(t)\big)$. Because this holds almost surely, simultaneously for all $t \in \mathbb{Q}$, and using right-continuity, we get that, almost surely, $(\pi_i(X(t)))_{t\geqslant 0}$ is the constant, null function. This implies that $\pi_i(\Delta X(t)) = 0$ for all $t \geqslant 0$, and thus $\pi_i(R_k) = 0$ for all $1 \leqslant k \leqslant d$. In other words, $\pi_i R = 0$, which implies $\lambda_i\pi_i = 0$, and thus $\lambda_i = 0$, which is impossible since, by assumption, $2\mathrm{Re}(\lambda_i) > \lambda > 0$. □

Finally, we look at the "small" eigenvalues

Lemma 1.11. *If $1 \leqslant i \leqslant d$ is such that $\mathrm{Re}(\lambda_i) < \lambda/2$, then $\pi_i(X(t)) = o(\mathrm{e}^{\lambda t})$ almost surely as $t \uparrow \infty$. Furthermore:*

- *If $1 \leqslant i \leqslant d$ is such that $\mathrm{Re}(\lambda_i) < \lambda/2$, then, in distribution as $t \uparrow \infty$,*
$$\mathrm{e}^{-\lambda t/2}\pi_i(X(t)) \Rightarrow \mathcal{N}(0, \sigma^2)v_i.$$
- *If $1 \leqslant i \leqslant d$ is such that $\mathrm{Re}(\lambda_i) = \lambda/2$, then, in distribution as $t \uparrow \infty$,*
$$t^{-1/2}\mathrm{e}^{-\lambda t/2}\pi_i(X(t)) \Rightarrow \mathcal{N}(0, \sigma^2)v_i.$$

This concludes the proof of Theorem 1.7. Indeed, we write

$$\mathrm{e}^{-\lambda t}X(t) = \sum_{i=1}^d \mathrm{e}^{-\lambda t}\pi_i(X(t))v_i.$$

for all $2 \leqslant i \leqslant d$, by Lemmas 1.9 and 1.11, $\mathrm{e}^{-\lambda t}\pi_i(X(t)) = o(1)$, and thus

$$\mathrm{e}^{-\lambda t}X(t) = \mathrm{e}^{-\lambda t}\pi_1(X(t))v + o(1) = Wv + o(1),$$

as desired.

1.2.4. *Proof of Theorem 1.5.* Theorem 1.5 follows from Theorem 1.7 through the connection made in Proposition 1.6. First note that Theorem 1.7 implies that $\mathrm{e}^{-\lambda t}\|X(t)\|_1 \to W$ almost surely, because $\|v\|_1 = 1$. This implies that

$$\frac{X(t)}{\|X(t)\|_1} \to v \quad \text{almost surely as } t \uparrow \infty. \tag{1.8}$$

Thus, by Proposition 1.6, and because $\tau_n \uparrow \infty$ almost surely as $n \uparrow \infty$,

$$\frac{U(n)}{\|U(n)\|_1} \to v \quad \text{almost surely as } n \uparrow \infty. \tag{1.9}$$

We now follow [AN72, V.7]. For all $1 \leqslant i \leqslant d$ and $k \geqslant 1$, let $\delta_k^{(i)}$ be the indicator that the k-th split in X is from a particle of colour i. For all $k \geqslant 1$, let $\mathcal{G}_k$ be the sub-σ-algebra generated by the process X until time τ_k. For all $k \geqslant 1$,

$$\mathbb{P}\big(\delta_k^{(i)} = 1|\mathcal{G}_k\big) = \frac{X_i(\tau_k^-)}{\|X(\tau_k^-)\|_1} \to v_i,$$

almost surely as $k \uparrow \infty$, by (1.8). For all $1 \leqslant i \leqslant d$ and $n \geqslant 0$, we let $N_i(n)$ be the number of splits triggered by a particle of colour i among the n first splits; for all $n \geqslant 0$,

$$\frac{N_i(n)}{n} = \frac{1}{n}\sum_{k=1}^n \delta_k^{(i)} = \frac{1}{n}\sum_{k=1}^n \mathbb{P}\big(\delta_k^{(i)} = 1|\mathcal{G}_k\big) + \frac{1}{n}\sum_{k=1}^n \big(\delta_k^{(i)} - \mathbb{P}\big(\delta_k^{(i)} = 1|\mathcal{G}_k\big)\big).$$

By a strong law of large numbers for martingales due to Lévy (see [AN72, V.7, Lemma 1]), the second summand goes to zero as $n \uparrow \infty$. By Cesáro's lemma, the first term converges almost surely to v_i. We thus get that, for all $1 \leqslant i \leqslant d$,

$$\frac{N_i(n)}{n} \to v_i \quad \text{almost surely as } n \uparrow \infty.$$

Now, for all $n \geqslant 0$,

$$\frac{\|X(\tau_n)\|_1}{n} = \frac{1}{n}\sum_{i=1}^d \sum_{j=1}^n \delta_j^{(i)} \|R_i\|_1 = \sum_{i=1}^d \|R_i\|_1 \frac{N_i(n)}{n} \to \sum_{i=1}^d v_i \|R_i\|_1,$$

almost surely as $n \uparrow \infty$. Now note that

$$\sum_{i=1}^d v_i\|R_i\|_1 = \sum_{i=1}^d v_i \sum_{j=1}^d R_{i,j} = \sum_{j=1}^d \sum_{i=1}^d v_i R_{i,j} = \sum_{j=1}^d (vR)_j = \|vR\|_1 = \|\lambda v\|_1 = \lambda.$$

Thus,

$$\frac{U(n)}{n} = \frac{\|X(\tau_n)\|_1}{n} \to \lambda \quad \text{almost surely as } n \uparrow \infty.$$

Together with (1.9), this concludes the proof of Theorem 1.5.

1.2.5. *A central limit theorem.* It is natural to ask for the fluctuations of $U(n)/n$ around its almost sure limit λv. Interestingly (although this should now feel natural in view of the proof of Theorem 1.5), the size and shape of these fluctuations depend on $\sigma = \max_\mu\{\mathrm{Re}(\mu)/\lambda\}$ where the maximum is taken over all eigenvalues of R except λ.

Theorem 1.12. *Under the assumptions of Theorem 1.5, let* $\sigma = \max_\mu\{\mathrm{Re}(\mu)/\lambda\}$ *where the maximum is taken over all eigenvalues of* R *except* λ.

(i) *If $\sigma < 1/2$, then, in distribution as $n \uparrow \infty$,*

$$\frac{U(n) - n\lambda v}{\sqrt{n}} \Rightarrow \mathcal{N}(0, \Sigma^2),$$

where Σ^2 is known and depends on R.

(ii) *If $\sigma = 1/2$, then let $\nu + 1$ denote the dimension of the largest generalised eigenspace of R associated to an eigenvalue of real part $\sigma\lambda$. With this notation, in distribution as $n \uparrow \infty$,*

$$\frac{U(n) - n\lambda v}{\sqrt{n(\log n)^{2\nu+1}}} \Rightarrow \mathcal{N}(0, \Sigma^2),$$

where Σ^2 is known and depends on R.

(iii) *If $\sigma > 1/2$, then let $\nu + 1$ denote the dimension of the largest generalised eigenspace of R associated to an eigenvalue of real part $\sigma\lambda$. With this notation, almost surely as $n \uparrow \infty$,*

$$\frac{U(n) - n\lambda v}{n^\sigma (\log n)^\nu} \text{ is tight.}$$

2. Infinitely-many colour Pólya urns

Infinitely-many-colour Pólya urns date back to Blackwell and McQueen [BM73] for the equivalent of the "identity" case of Section 1. For the equivalent of the "irreducible" case of Section 1, they are much more recent and only date back to Bandyopadhyay and Thacker [BT16] and Mailler and Marckert [MM17], with a particular case dating back to 2013 (see [BT17]). In this course, we call infinitely-many-colour Pólya urns "measure valued Pólya processes" (MVPPs), and give their definition in general, before looking at results for the "identity" case on the one hand, and the "irreducible" case on the other hand.

2.1. Definition of MVPPs.

Definition. Let E be a Polish space, m_0 be a finite measure on E and let $R^{(1)} = (R^{(1)}_x)_{x\in E}$ be a random kernel on E, i.e., for all $x \in E$, $R^{(1)}_x(E)$ is a finite measure on E. The measure-valued Pólya process (MVPP) of initial composition m_0 and replacement kernel $R^{(1)}$ is defined recursively as follows: for all $n \geqslant 0$, given m_n, let

$$m_{n+1} = m_n + R^{(n+1)}_{Y(n+1)},$$

where $Y(n+1)$ is a random variable of distribution $m_n/m_n(E)$ and, given $Y(n+1)$, $R^{(n+1)}$ is an independent copy of $R^{(1)}_{Y(n+1)}$.

In the language of Pólya urns, E is the colour space; for all Borel sets $B \subset E$ and integers $n \geqslant 0$, $m_n(B)$ is the mass of balls in the urn at time n whose colour belong to B; for all $n \geqslant 1$, $Y(n)$ is the colour of the ball drawn at time n; note that balls are not necessarily of mass 1 anymore, and they can in fact have infinitesimal mass.

Example: the finitely-many-colour case. Take $E = \{1, \dots, d\}$ for some integer $d \geqslant 2$. Consider a matrix $\mathfrak{r} = (\mathfrak{r}_{i,j})_{1\leqslant i,j\leqslant d}$ and a vector $\alpha = (\alpha_1, \dots, \alpha_d)$ both integer-valued. Let $(m_n)_{n\geqslant 0}$ be the MVPP of initial composition $m_0 = \sum_{i=1}^d \alpha_i\delta_i$ (where δ_i is the Dirac measure at $\{i\}$), and deterministic replacement kernel $R_i = \sum_{j=1}^d \mathfrak{r}_{i,j}\delta_j$. Then, in distribution, $(m_n = \sum_{i=1}^d U_i(n)\delta_i)_{n\geqslant 0}$ where $(U(n))_{n\geqslant 0}$ is the Pólya urn of initial composition α and replacement matrix $\mathfrak{r}$. In other words, all finitely-many-colour Pólya urns are MVPPs.

2.2. **The "identity" case.** In the following theorem, Blackwell and McQueen [BM73] show that, in the "identity" case, which corresponds to $R_x = \delta_x$ for all $x \in E$, the composition measure m_n converge after renormalisation to a limiting random measure, whose distribution is the Ferguson distribution, which we define now:

Definition. Let μ be a finite measure on a Polish space E. A random measure on E has Ferguson distribution of parameter μ if for every Borel set partition $B_1, \ldots, B_d$ of E ($d \geqslant 1$), $(\mu^*(B_1), \ldots, \mu^*(B_d))$ is Dirichlet-distributed with parameter $(\mu(B_1), \ldots, \mu(B_d))$.

Ferguson [Fer73] proved that a Ferguson-distributed random measure is almost surely discrete.

Theorem 2.1. *Let μ be a finite measure on a Polish space E. Let $(m_n)_{n\geqslant 0}$ be the MVPP of initial composition μ and deterministic replacement kernel* $\mathrm{Id} = (\delta_x)_{x\in E}$. *Then, almost surely as $n \uparrow \infty$,*

$$\hat{m}_n := \frac{m_n}{m_n(E)} \to \mu^*,$$

where μ^* is a Ferguson-distributed with parameter μ.

NB: In this section, we often state convergence of sequences of measures on a Polish space E. The topology we use on the set of finite measures on E is the topology of weak convergence. I.e., we say the a sequence $(\mu_n)_{n\geqslant 0}$ of finite measures on E converges to μ if, and only if, for all continuous and bounded function $f : E \to \mathbb{R}$, $\int_E f \mathrm{d}\mu_n \to \int_E f \mathrm{d}\mu$.

Proof. One can redefine the process $(m_n)_{n\geqslant 0}$ as follows: $m_0 = \mu$ and, for all $n \geqslant 0$, given m_n, sample an integer in $\{0, 1, \ldots, n\}$ with probabilities

$$\mathbb{P}(I(n+1) = 0) = \frac{\mu(E)}{\mu(E) + n} \quad \text{ and } \mathbb{P}(I(n+1) = i) = \frac{1}{\mu(E) + n} (\forall 1 \leqslant i \leqslant n).$$

If $I(n+1) = 0$, then sample $Y(n+1)$ according to μ. If $I(n) = i$ for some $1 \leqslant i \leqslant n$, then set $Y(n+1) = Y(i)$. One can check that, indeed, $(m_n)_{n\geqslant 0}$ is the MVPP of initial composition μ and replacement kernel Id.

We see the times n such that $I(n) = 0$ as times at which we sampled a "new" colour from μ, while other times reinforced a colour that was already present in the urn. (Note that, if μ has atoms, then a new colour might in fact be the same as an already-present colour. In that case, we still distinguish the new colour from the existing colour, eg, by extending the set of colours to $E \times \mathbb{N}$ and deciding that when colour $x \in E$ is picked for the second time, it is in fact colour $(x, 2)$ while its first occurrence was $(x, 1)$. However, to make things simpler, we assume that ν has no atoms in the rest of the proof.) We let ν_i be the index of the i-th 0 in the sequence $(i(n))_{n\geqslant 1}$ and set $\xi_i = Y(\nu_i)$, for all $i \geqslant 1$. By definition, $(\xi_i)_{i\geqslant 1}$ is a sequence of i.i.d. random variables of distribution μ.

We start by looking at $N_1(n) = |\{1 \leqslant i \leqslant n \colon Y(i) = Y(1)\}|$. Note that $(N_1(n), n + \mu(E) - N_1(n))_{n\geqslant 1}$ is a Pólya urn of initial composition $(1, \mu(E))$ and replacement matrix Id. (NB: $\mu(E)$ might not be an integer, but one can check that all the results from Section 1 still apply.) Thus, by Corollary 1.2, almost surely as $n \uparrow \infty$,

$$\frac{N_1(n)}{\mu(E) + n} \to V_1,$$

where V_1 is a Beta-distributed random variable of parameter $(1, \mu(E))$. This implies

$$\frac{N_1(n)}{n} \to V_1$$

almost surely as $n \uparrow \infty$. By Borel-Cantelli lemma, $\sum_{n\geqslant 1} \mathbf{1}_{I(n)=0} = \infty$ almost surely, implying in particular that the time τ_2 at which a second colour is introduced in the urn is almost surely finite.

We let $N_2(n) = |\{1 \leqslant i \leqslant n \colon Y(i) = \xi_2\}|$. Note that the processus $(N_2(n), n + \mu(E) - N_1(n) - N_2(n))_{n\geqslant 0}$ reduced to its jump-times is also a Pólya urn, with initial composition $(1, \mu(E))$ and replacement matrix Id. Therefore,

$$\frac{N_2(n)}{n + \mu(E) - N_1(n)} \to V_2,$$

where V_2 is a Beta-distributed random variable of parameter $(1, \mu(E))$, independent of V_1. Thus,

$$\frac{N_2(n)}{n} = \frac{N_2(n)}{n + \mu(E) - N_1(n)} \frac{n + \mu(E) - N_1(n)}{n} \to V_2(1 - V_1).$$

Iterating this argument, we get that, for all $k \geqslant 1$, almost surely as $n \uparrow \infty$,

$$\frac{N_k(n)}{n} \to V_k \prod_{i=1}^{k-1}(1 - V_i) =: P_k.$$

Such a sequence $(P_k)_{k\geqslant 1}$ can be interpreted as a partition of the unit interval and, in this context, because $(V_i)_{i\geqslant 1}$ is a sequence of i.i.d. random variables of distribution Beta of parameter $(1, \mu(E))$, the distribution $(P_k)_{k\geqslant 1}$ is called GEM (for Griffiths, Engen and McCloskey) of parameter $\mu(E)$. Now note that, for all $n \geqslant 1$,

$$m_n = \mu + \sum_{k\geqslant 1} N_k(n)\delta_{\xi_k},$$

where we set $N_k(n) = 0$ if $|\{Y(1), \dots, Y(n)\}| < k$. Thus, for all continuous bounded function $f : E \to \mathbb{R}$,

$$\int_E f \mathrm{d}\hat{m}_n = \frac{1}{n + \mu(E)} \int_E f \mathrm{d}\mu + \frac{n}{n + \mu(E)} \sum_{k\geqslant 1} \frac{N_k(n)}{n} f(\xi_k).$$

Fix $\varepsilon > 0$ and then K large enough so that $\sum_{k=1}^K P_k \geqslant 1 - \varepsilon$ (this is possible because $\sum_{k\geqslant 1} P_k = 1$ almost surely). Then fix n_ε large enough so that, for all $n \geqslant n_\varepsilon$,

$$\left| \sum_{k=1}^K \Big(\frac{N_k(n)}{n} - P_k \Big) \right| \leqslant \varepsilon. \tag{2.1}$$

We also assume that n_ε is large enough so that $(\int_E f \mathrm{d}\mu)/(n + \mu(E)) < \varepsilon$ for all $n \geqslant n_\varepsilon$. Then, for all $n \geqslant n_\varepsilon$,

$$\begin{aligned}
\left| \int_E f \mathrm{d}\hat{m}_n - \sum_{k\geqslant 1} P_k f(\xi_k) \right| &\leqslant \varepsilon + \frac{n}{n + \mu(E)} \left| \sum_{k=1}^K \Big(\frac{N_k(n)}{n} - P_k \Big) f(\xi_k) + \sum_{k\geqslant K} \Big(\frac{N_k(n)}{n} - P_k \Big) f(\xi_k) \right| \\
&\leqslant \varepsilon + \sum_{k=1}^K \left| \frac{N_k(n)}{n} - P_k \right| \|f\|_\infty + \sum_{k\geqslant K} \left| \frac{N_k(n)}{n} - P_k \right| \|f\|_\infty \\
&\leqslant (1 + \|f\|_\infty)\varepsilon + \|f\|_\infty \sum_{k\geqslant K} \frac{N_k(n)}{n} + \|f\|_\infty \sum_{k\geqslant K} P_k \\
&\leqslant (1 + 2\|f\|_\infty)\varepsilon + \|f\|_\infty \Bigg(1 - \sum_{i=1}^K \frac{N_k(n)}{n} \Bigg).
\end{aligned}$$

Now,

$$\sum_{i=1}^K \frac{N_k(n)}{n} = \sum_{i=1}^K \Big(\frac{N_k(n)}{n} - P_k \Big) + \sum_{i=1}^K P_k \geqslant - \sum_{i=1}^K \left| \frac{N_k(n)}{n} - P_k \right| + 1 - \varepsilon \geqslant 1 - 2\varepsilon.$$

In total, we thus get that, for all $n \geqslant n_\varepsilon$,

$$\left| \int_E f \mathrm{d}\hat{m}_n - \sum_{k\geqslant 1} P_k f(\xi_k) \right| \leqslant (1 + 4\|f\|_\infty)\varepsilon.$$

Because ε can be made arbitrarily small, this concludes the proof that

$$\int_E f \mathrm{d}\hat{m}_n \to \sum_{k\geqslant 1} P_k f(\xi_k),$$

almost surely as $n \uparrow \infty$. We define

$$\mu^* = \sum_{k\geqslant 1} P_k \delta_{\xi_k}.$$

It only remains to prove that μ^* is a Ferguson-distributed with parameter μ. This follows from the finitely-many-colour case (see Theorem 1.1). Indeed, for any partition $(B_1, \ldots, B_d)$ of E, we let, for all $i \geqslant 1$,

$$X(i) = \sum_{k=1}^{d} e_k \mathbf{1}_{Y(i)\in B_k},$$

i.e. $X(i) = e_k$ (the k-th vector of the canonical basis of $\mathbb{R}^d$) if $Y(i) \in B_k$. For all $n \geqslant 0$, we define

$$U(n) = \sum_{k=1}^{d} \mu(B_k)e_k + \sum_{i=1}^{n} e_{X(i)}.$$

Note that, for all $n \geqslant 0$,

$$U(n+1) = U(n) + e_{X(n+1)},$$

and

$$\begin{aligned}
\mathbb{P}(X(n+1) = k|U(n)) &= \mathbb{P}(Y(n+1) \in B_k|U(n)) = \mathbb{E}[\mathbb{P}(Y(n+1) \in B_k|U(n), m_n)|U(n)] \\
&= \mathbb{E}[\hat{m}_n(B_k)|U(n)] = \mathbb{E}\left[\frac{\mu(B_k) + \sum_{i=1}^n \mathbf{1}_{Y(i)\in B_k}}{\mu(E) + n}\middle|U(n)\right] \\
&= \mathbb{E}\left[\frac{\mu(B_k) + \sum_{i=1}^n \mathbf{1}_{Y(i)\in B_k}}{\mu(E) + n}\middle|U(n)\right] = \frac{U_k(n)}{\mu(E) + n}.
\end{aligned}$$

In other words, $(U(n))_{n\geqslant 0}$ is the d-colour Pólya urn with initial composition $(\mu(B_1), \ldots, \mu(B_d))$ and replacement matrix Id. By Theorem 1.1), the almost sure limit of $U(n)/n$, which we know is $(\mu^*(B_1), \ldots, \mu^*(B_n))$ is Dirichlet-distributed of parameter $(\mu(B_1), \ldots, \mu(B_d))$. This concludes the proof that μ^* is indeed Ferguson-distributed with parameter μ. ☐

However, Blackwell and McQueen [BM73] also prove that Theorem 1.3 still holds in this case:

Theorem 2.2. *Let μ be a finite measure on a Polish space E and let μ^* be a Ferguson(μ)-distributed random probability measure on E. Given μ^*, let $(X(i))_{i\geqslant 1}$ be a sequence of i.i.d. random variables of distribution μ^*, and let, for all $n \geqslant 0$,*

$$m_n = \mu + \sum_{i=1}^{n} \delta_{X(i)}.$$

Then, $(m_n)_{n\geqslant 0}$ is the MVPP of initial composition μ and replacement kernel Id.

Proof. This can be done using de Finetti's theorem. ☐

For the fluctuations of $\hat{m}_n$ around its almost sure limit μ^*, one can use Theorem 2.2. See Borovkov [**?**] for a functional limit theorem for these fluctuations.

2.3. **The balanced "irreducible" case.** In this section, which is based on Mailler and Marckert [MM17], we assume that the MVPP is balanced, i.e. that there exists $c > 0$ such that, for all $x \in E$, $R_x(E) = 1$. Without loss of generality, we assume that $c = 1$. In other words, for all $x \in E$, R_x is a probability measure; we call R a "probability kernel". The non-balanced case is treated in Section 2.4.

2.3.1. *Convergence in probability of "irreducible" MVPPs.*

Definition. For any two E-valued sequences $a = (a(n))_{n\geqslant 0}, b = (b(n))_{n\geqslant 0}$, and any probability distribution ν on E, we say that probability kernel $(R_x)_{x\in E}$ on E is "(a, b, ν)-ergodic" if, for all $x \in E$, the Markov chain $(W(n))_{n\geqslant 0}$ started at x and whose transition probabilities are given by R (i.e. $W(n+1) \sim R_{W(n)}$, for all $n \geqslant 0$) satisfies

$$\frac{W(n) - a(n)}{b(n)} \Rightarrow \nu,$$

in distribution as $n \uparrow \infty$.

NB: Note that we have not assumed that E is equipped with an addition and a multiplication operation. If there is no addition on E, then the only possible value for $(a(n))_{n\geqslant 0}$ is the constant sequence equal to 0, and we interpret $x \mapsto x + 0$ as being the identity function on E. Similarly, if there is no multiplication by a scalar on E, then the only possible value for $(b(n))_{n\geqslant 0}$ is the constant sequence equal to 1, and we interpret $x \mapsto 1 \times x$ as being the identity function on E.

Theorem 2.3. *Let $(m_n)_{n\geqslant 0}$ be the MVPP of initial composition m_0 and balanced replacement kernel $(R_x)_{x\in E}$. We assume that there exist two sequences $a = (a(n))_{n\geqslant 0}, b = (b(n))_{n\geqslant 0}$, and a probability measure ν such that $(R_x)_{x\in E}$ is (a, b, ν)-ergodic and there exist $f, g : E \to \mathbb{R}$ such that, for all $x \in E$, for any function $(\varepsilon_n)_{n\geqslant 0}$ satisfying $\varepsilon_n = o(\sqrt{n})$ as $n \uparrow \infty$,*

$$\frac{a(n + x\sqrt{n} + \varepsilon_n) - a(n)}{b(n)} \to f(x), \quad \text{and} \quad \frac{b(n + x\sqrt{n} + \varepsilon_n)}{b(n)} \to g(x), \tag{2.2}$$

as $n \uparrow \infty$. Recall that, for all $n \geqslant 0$, $\hat{m}_n = m_n/m_n(E)$. Then, in probability as $n \uparrow \infty$,

$$\hat{m}_n(a(\log n) + \cdot\, b(\log n)) \to \mathcal{L}(f(\Omega) + g(\Omega)\Gamma), \tag{2.3}$$

where $\Omega \sim \mathcal{N}(0, 1)$ and $\Gamma \sim \nu$ are independent, and, for any random variable X, $\mathcal{L}(X)$ denotes its law.

NB: Equation (2.3) means that, for any continuous bounded function $\varphi : E \to \mathbb{R}$,

$$\int_E \varphi\left(\frac{x - a(\log n)}{b(\log n)}\right) \mathrm{d}\hat{m}_n(x) \to \mathbb{E}\big[\varphi(f(\Omega) + g(\Omega)\Gamma)\big],$$

in probability as $n \uparrow \infty$.

The proof of Theorem 2.3 can be summarised in one sentence: "*The Pólya urn of replacement matrix R is the branching Markov chain of transition probabilities R indexed by the random recursive tree.*"

Before proving Theorem 2.3, we give some examples of applications.

2.3.2. *Examples of application of Theorem 2.3.* We give two examples of applications of Theorem 2.3.

(1) Finitely-many colour Pólya urns: Let $(U(n))_{n\geqslant 0}$ be the d-colour Pólya urn of replacement matrix $\mathfrak{r} = (\mathfrak{r}_{i,j})_{1\leqslant i,j\leqslant d}$ and initial composition $\alpha = (\alpha_1, \ldots, \alpha_d)$. Recall that $(m_n = \sum_{i=1}^d U(n)\delta_i)_{n\geqslant 0}$ is the MVPP of initial composition $m_0 = \sum_{i=1}^d \alpha_i\delta_i$ and replacement kernel $(R_i = \sum_{j=1}^d \mathfrak{r}_{i,j}\delta_j)_{1\leqslant i\leqslant d}$.

To apply Theorem 2.3, we need to assume that, for all $1 \leqslant i \leqslant d$, $\sum_{j=1}^d \mathfrak{r}_{i,j} = \mathfrak{s}$ for some $\mathfrak{s} \geqslant 1$; this assumption is classical in the literature and is called the "balance" assumption. For all $n \geqslant 0$, we let $\eta_n = m_n/\mathfrak{s}$; $(\eta_n)_{n\geqslant 0}$ is a balanced MVPP of replacement kernel $R/\mathfrak{s}$ and initial composition $\alpha/\mathfrak{s}$.

If, in addition, we assume that $\mathfrak{r}$ is irreducible and aperiodic, then the Markov chain of transition probabilities $R/\mathfrak{s}$ is ergodic. Furthermore, its limiting probability distribution satisfies $v(R/\mathfrak{s}) = v$, which is equivalent to $vR = \mathfrak{s}v$. Theorem 2.3 thus applies and gives implies that

$$\frac{U(n)}{\|U(0)\|_1 + n\mathfrak{s}} \to v \quad \text{ in probability as } n \uparrow \infty.$$

Because R is balanced and all its rows sum to $\mathfrak{s}$, $\mathfrak{s}$ is its dominant eigenvalue, and thus v is its dominant left-eigenvector. Thus, we recover a version of Theorem 1.5. This version however is weaker, for three reasons: first, we assumed that the Pólya urn is balanced; second, we assumes that R is aperiodic, and third, we get convergence in probability instead of almost surely.

(2) The "random walk increment" replacement kernel. Let $(m_n)_{n\geqslant 0}$ be the MVPP of initial composition m_0 and replacement kernel $(\mathcal{R}_x)_{x\in E}$, where, for all $x \in \mathbb{R}^d$, R_x is the probability distribution of $x + \Delta$ with Δ a random variable on $\mathbb{R}^d$ such that $\mu = \mathbb{E}\Delta$ and $\sigma^2 = \mathrm{Var}(\Delta) < \infty$. Then, the Markov chain of probability transitions $(R_x)_{x\in\mathbb{R}^d}$ is the simple random walk of increment Δ. I.e., for all $n \geqslant 0$, $W_n = W_0 + \sum_{i=1}^n \Delta_i$, where $(\Delta_i)_{i\geqslant 1}$ is a sequence of i.i.d. copies of Δ. By the central limit theorem, $(W_n)_{n\geqslant 0}$ is (a, b, ν)-ergodic with $a(n) = mn$, $b(n) = \sqrt{n}$ and $\nu = \mathcal{N}(0, \sigma^2)$. Note that a and b satisfy (2.2) with $g(x) = 1$ and $f(x) = mx$. Theorem 2.3 thus implies that, in probability as $n \uparrow \infty$,

$$\frac{m_n}{m_n(E)} \to \mathcal{L}(m\Omega + \Gamma),$$

where $\Omega \sim \mathcal{N}(0, 1)$ and $\Gamma \sim \mathcal{N}(0, \sigma^2)$ are independent. In other words, in probability as $n \uparrow \infty$,

$$\frac{m_n}{m_n(E)} \to \mathcal{N}(0, m^2 + \sigma^2).$$

2.3.3. *MVPPs are branching Markov chains.* This section is a preliminary to the proof of Theorem 2.3, we start by giving the following, equivalent, definition of an MVPP: given a replacement kernel R and an initial composition m_0, we define the sequence $(m_n)_{n\geqslant 0}$ as follows: first sample a sequence $(I(n))_{n\geqslant 1}$ of independent random variables such that, for all $n \geqslant 1$,

$$\mathbb{P}(I(n) = 0) = \frac{m_0(E)}{m_0(E) + n - 1} \quad \text{ and } \quad \mathbb{P}(I(n) = i) = \frac{1}{m_0(E) + n - 1} \ (1 \leqslant i \leqslant n - 1).$$

Then, for all $n \geqslant 0$, given $Y(1), \dots, Y(n)$ (recall that $Y(i)$ is the colour of the ball drawn at time i), if $I(n + 1) = i$ for some $1 \leqslant i \leqslant n$, then sample $Y(n + 1)$ according to the probability distribution $R_{Y(i)}$, otherwise (if $I(n + 1) = 0$), sample $Y(n + 1)$ according to m_0. Finally, set $m_n = m_0 + \sum_{i=1}^n R_{Y(i)}$ for all $n \geqslant 0$. One can check that, indeed, $(m_n)_{n\geqslant 0}$ is the MVPP of replacement kernel R and an initial composition m_0.

The advantage of this definition is that the sequence $(I(n))_{n\geqslant 1}$ records the branching structure of the MVPP: indeed, we can interpret $I(n)$ as the parent of n, for all $n \geqslant 1$. With this definitions, all positive integers have a parent, and the genealogical structure of this population can be interpreted as a tree. To do so, we just interpret each integer as a node, and add an edge between every node and its parent (because one's parent is a smaller integer, there are indeed no cycles in this graph). The only node that has no parent is 0, and we call this node the root of the tree. The tree T_∞ associated to $(I(n))_{n\geqslant 0}$ is infinite; for all $n \geqslant 1$, we call T_n the tree associated to the finite sequence $(I(1), \dots, I(n))$. Because $(I(n))_{n\geqslant 0}$ is a random sequence, $(T_n)_{n\geqslant 0}$ is a sequence of random trees.

If $m_0(E) = 1$ (and, for simplicity, we now assume that this is the case), then $(T_n)_{n\geqslant 0}$ is known in the literature as the random recursive tree (RRT). Now, the sequence $(Y(i))_{i\geqslant 1}$ of the colours drawn at successive times in the MVPP can be interpreted, not as a sequence indexed by $\mathbb{N}$, but as a sequence indexed by the RRT T_∞. From now on, we call $Y(i)$ the “label” of node i. The sequence of labels taken along each branch of the tree is distributed as a Markov chain of initial distribution m_0 and transition probabilities R. And once they have branched, the labels along two distinct branches are two independent Markov chains of transition probabilities R. In other words, $(Y(i))_{i\geqslant 0}$ seen as a sequence indexed by T_∞ is a branching Markov chain.

The advantage of this description if the following result:

Lemma 2.4. *Let R be a replacement kernel on a Polish space E and m_0 a finite measure on E. We assume that $R_x(E) = 1$ for all $x \in E$. Let T_∞ be the random recursive tree whose nodes are $\{0, 1, 2, \dots\}$, and for all $n \geqslant 0$, let T_n be the restriction of T_∞ to $\{0, \dots, n\}$. Let $(Y(i))_{i\geqslant 1}$ be the branching Markov chain of initial distribution m_0 and transition probabilities R indexed by T_∞. Then, for all $n \geqslant 0$, given T_n and $(Y(i))_{1\leqslant i\leqslant n}$, $\hat{m}_n = m_n/m_n(E)$ is the distribution of $Y(n+1)$.*

Thanks to this coupling, we are now ready to prove Theorem 2.3.

2.3.4. *Proof of Theorem 2.3.* For this proof, we “forget” about this MVPP and now focus on the branching Markov chain on the random recursive tree. First note that, in distribution, $Y(n+1) = W_{|n+1|}$, where $|n+1|$ denotes the height of node $n+1$ in T_∞ and W is the Markov chain of initial distribution m_0 and replacement probabilities R. Our assumptions give us good control of $W_{|n+1|}$ under the assumption that $|n+1|$ tends to infinity. The following result gives us good control of $|n+1|$ as n tends to infinity; in fact, instead of looking at node $n+1$, we look at its (random) parent $\xi(n+1)$. Note that $\xi(n+1)$ is uniformly distributed in $\{0, \dots, n\}$.

Lemma 2.5 (Dobrow [Dob96])**.** *Let T_∞ be therandom recursive tree. Let $\xi(n+1)$ be a node taken uniformly at random among $\{0, 1, \dots, n+1\}$. Then, the height of $\xi(n+1)$, which $|\xi(n+1)|$ denotes, satisfies, in distribution as $n \uparrow \infty$,*

$$\frac{|\xi(n+1)| - \log n}{\sqrt{\log n}} \Rightarrow \mathcal{N}(0,1). \tag{2.4}$$

Recall that, by assumption, $(W_n - a(n))/b(n) \Rightarrow \nu$. By Kolmogorov’s representation theorem, there exists a probability space, which we call “Kolmogorov’s space” in the following, in which this convergence and the convergence of (2.4) hold simultaneously almost surely: (for simplicity, we keep the same notation, although we are in fact working with different objects, on a different probability space)

$$\left(\frac{|\xi(n+1)| - \log n}{\sqrt{\log n}}, \frac{W_n - a(n)}{b(n)}\right) \to (\Omega, \Gamma),$$

almost surely as $n \uparrow \infty$, where $\Omega \sim \mathcal{N}(0,1)$ and $\Gamma \sim \nu$ are independent. Now,

$$\begin{aligned}\frac{W_{|n+1|} - a(\log n)}{b(\log n)} &= \frac{W_{|\xi(n+1)|+1} - a(|\xi(n+1)|+1)}{b(|\xi(n+1)|+1)} \cdot \frac{b(|\xi(n+1)|+1)}{b(\log n)} + \frac{a(|\xi(n+1)|+1) - a(\log n)}{b(\log n)} \\ &\to \Gamma g(\Omega) + f(\Omega), \end{aligned} \tag{2.5}$$

almost surely as $n \uparrow \infty$. The almost sure convergence is only true on Kolmogorov’s probability space; however, this implies convergence in distribution on the original probability space. We thus have shown that

$$\frac{W_{|n+1|} - a(\log n)}{b(\log n)} \Rightarrow \mathcal{L}(\Gamma g(\Omega) + f(\Omega)), \tag{2.6}$$

in distribution as $n \uparrow \infty$. Equivalently, for any bounded continuous function $\varphi : E \mapsto R$,

$$\mathbb{E}\bigg[\varphi\bigg(\frac{W_{|n+1|} - a(\log n)}{b(\log n)}\bigg)\bigg] \to \mathbb{E}[\varphi(\Gamma g(\Omega) + f(\Omega))].$$

Unfortunately, even if $(W_{|n+1|} - a(\log n))/b(\log n)$ is a random variable of distribution $\hat{m}_n(a(\log n) + \cdot\, b(\log n))$, this is not enough to prove convergence of $\hat{m}_n(a(\log n) + \cdot\, b(\log n))$ to $\mathcal{L}(\Gamma g(\Omega) + f(\Omega))$ in probability as $n \uparrow \infty$. Indeed, to do so, we need more, as stated in the following result (which is folklore, and proved in [MM17]):

Lemma 2.6. *Let $(\mu_n)_{n\geqslant 0}$ be a sequence of random probability distributions on a Polish space E. Let μ be a (deterministic) probability distribution on E. For all $n \geqslant 0$, given m_n, let A_n and B_n be two independent random variables of distribution μ_n. If $(A_n, B_n) \Rightarrow (A, B)$ in distribution as $n \uparrow \infty$, where A and B are two independent random variables of distribution μ, then $\mu_n \to \mu$ in probability for the weak topology. I.e., for any bounded continuous function $\varphi : E \mapsto R$, in probability as $n \uparrow \infty$,*

$$\int_E \varphi \,\mathrm{d}\mu_n \to \int_E \varphi \,\mathrm{d}\mu.$$

Proof. By Markov's inequality, because μ is deterministic, it is enough to prove that, for any bounded continuous function $\varphi : E \mapsto R$, as $n \uparrow \infty$,

$$\mathbb{E}\bigg[\int_E \varphi \,\mathrm{d}\mu_n\bigg] \to \int_E \varphi \,\mathrm{d}\mu \quad \text{and} \quad \mathrm{Var}\bigg(\int_E \varphi \,\mathrm{d}\mu_n\bigg) \to 0. \tag{2.7}$$

First, because, by definition of A_n, $\int_E \varphi \,\mathrm{d}\mu_n = \mathbb{E}[\varphi(A_n)|\mu_n]$,

$$\mathbb{E}\bigg[\int_E \varphi \,\mathrm{d}\mu_n\bigg] = \mathbb{E}[\mathbb{E}[\varphi(A_n)|\mu_n]] = \mathbb{E}[\varphi(A_n)] \to \mathbb{E}[\varphi(A)] = \int_E \varphi \,\mathrm{d}\mu, \tag{2.8}$$

since $A_n \Rightarrow A$ in distribution as $n \uparrow \infty$. Also, because $\mathbb{E}[A_n|\mu_n] = \mathbb{E}[B_n|\mu_n] = \int_E \varphi \,\mathrm{d}\mu_n$, we have

$$\mathbb{E}\bigg[\bigg(\int_E \varphi \,\mathrm{d}\mu_n\bigg)^2\bigg] = \mathbb{E}\big[\mathbb{E}[A_n|\mu_n]\mathbb{E}[B_n|\mu_n]\big] = \mathbb{E}[\mathbb{E}[A_nB_n|\mu_n]],$$

because, given m_n, A_n and B_n are independent. Thus,

$$\mathbb{E}\bigg[\bigg(\int_E \varphi \,\mathrm{d}\mu_n\bigg)^2\bigg] = \mathbb{E}[A_nB_n] \to \mathbb{E}[AB] = \mathbb{E}[A]\mathbb{E}[B] = \bigg(\int_E \varphi \,\mathrm{d}\mu\bigg)^2.$$

Together with (2.8), this implies (2.7) and thus concludes the proof. □

Thus, to prove Theorem 2.3, it is enough to prove the following:

Proposition 2.7. *Let R be a balanced replacement kernel on a Polish space E and m_0 be a probability measure on E. Let T_∞ be the random recursive tree and let $(Y(i))_{i\geqslant 1}$ be the branching Markov chain of initial distribution m_0 and transition probabilities R indexed by T_∞. Given T_n and $(Y(i))_{1\leqslant i\leqslant n}$, let $Y_1(n+1)$ and $Y_2(n+1)$ two independent copies of $Y(n+1)$. Then, in distribution as $n \uparrow \infty$,*

$$\bigg(\frac{Y_1(n+1) - a(\log n)}{b(\log n)}, \frac{Y_2(n+1) - a(\log n)}{b(\log n)}\bigg) \Rightarrow \big(f(\Omega_1) + g(\Omega_1)\Gamma_1, f(\Omega_2) + g(\Omega_2)\Gamma_2\big),$$

where $\Omega_1, \Omega_2 \sim \mathcal{N}(0, 1)$ and $\Gamma_1, \Gamma_2 \sim \nu$ are all independent.

To prove this proposition, we neet to improve Dobrow's lemma (Lemma 2.5) as follows:

Lemma 2.8. *Let T_∞ be the random recursive tree. Let $\zeta(n+1)$ and $\xi(n+1)$ be two nodes taken independently uniformly at random in $\{0,1,\ldots,n\}$. Let $|\zeta(n+1)|$, $|\xi(n+1)|$ be their respective heights in T_∞, and let K_n be the height of their last common ancestor in T_∞. Then, in distribution as $n\uparrow\infty$,*

$$\left(\frac{|\zeta(n+1)|-\log n}{\sqrt{\log n}}, \frac{|\xi(n+1)|-\log n}{\sqrt{\log n}}, K_n\right) \Rightarrow (\Omega_1,\Omega_2,K), \tag{2.9}$$

where Ω_1 and Ω_2 are two independent standard Gaussian random variables, and K an almost surely finite random variable.

This lemma is stated in [MM17] but with a different proof. There is also a version of this result in [MB19] but with a proof that is different from that of [MM17] and different from the one that follows; the proof of [MB19] is an extension of Dobrow's proof of Lemma 2.5.

Proof. Our proof relies on a coupling that couples all the triples $(T_n, \xi(n+1), \zeta(n+1))$ for $n\geqslant 0$ in such a way that, for all $n\geqslant 1$,

- $\zeta(n+1)$ is either equal to $\zeta(n)$ or a child of $\zeta(n)$, and
- $\xi(n+1)$ is either equal to $\xi(n+1)$, or equal to $\xi(n)$, or a child of $\xi(n)$.

The coupling goes as follows: first let $\xi(1)=\zeta(1)=0$. Then sample two independent sequences of independent random variables $(B_i)_{i\geqslant 1}$ and $(B_i')_{i\geqslant 1}$ such that, for all $i\geqslant 1$,

$$\mathbb{P}(B_i=1)=\mathbb{P}(B_i'=1)=\frac{1}{i} \quad\text{and}\quad \mathbb{P}(B_i=0)=\mathbb{P}(B_i'=0)=1-\frac{1}{i}.$$

Also sample a sequence $(U(n))_{n\geqslant 1}$ of independent random variables such that, for all $n\geqslant 1$, $U(n)$ is uniform on $\{0,\ldots,n-1\}$. Then, assuming that $(T_{n-1},\zeta(n),\xi(n))$ is defined, we define $(T_n,\zeta(n+1),\xi(n+1))$ as follows:

- If $B_n=B_n'=0$, then set $\zeta(n+1)=\zeta(n)$, $\xi(n+1)=\xi(n)$, and T_n is the tree obtained after adding node n in T_{n-1} as a child of node $U(n)$.
- If $B_n=B_n'=1$, then set $\zeta(n+1)=\xi(n+1)=n+1$, and let T_n be the tree obtained after adding node n to T_{n-1} as a child of node $\zeta(n)$.
- If $B_n=1$ and $B_n'=0$, then set $\zeta(n+1)=n+1$, $\xi(n+1)=\xi(n)$, and let T_n be the tree obtained after adding node n to T_{n-1} as a child of node $\zeta(n)$.
- If $B_n=0$ and $B_n'=1$, then set $\zeta(n+1)=\zeta(n)$, $\xi(n+1)=n+1$, and let T_n be the tree obtained after adding node n to T_{n-1} as a child of node $\xi(n)$.

One can check that, indeed, $(T_n)_{n\geqslant 0}$ is distributed as the random recursive tree, and for all $n\geqslant 0$, $\zeta(n+1)$ and $\xi(n+1)$ are two nodes taken uniformly at random among $\{0,\ldots,n\}$. With this definition, we have the following identities: for all $n\geqslant 0$,

$$|\zeta(n+1)|=\sum_{i=1}^n B_i,\quad K_n=\sum_{i=1}^{I_n} B_i,\quad\text{and}\quad |\xi(n+1)|=K_n+\sum_{i=I_n+1}^n B_i',$$

where $I_n=\max\{1\leqslant i\leqslant n\colon B_i=B_i'=1\}$. First note that, by Borel-Cantelli's lemma, because $\mathbb{P}(B_i=B_i'=1)={}^1\!/_{i^2}$, $I=\lim_{n\uparrow\infty} I_n=\sum_{i\geqslant 1} B_iB_i'<\infty$ almost surely. This implies that, $K=\lim_{n\uparrow\infty}K_n<\infty$ almost surely. Furthermore, by the central limit theorem, because

$$\sum_{i=1}^n \mathbb{E}B_i=\sum_{i=1}^n\frac{1}{i}=\log n+\mathcal{O}(1),$$

and

$$\sum_{i=1}^n \operatorname{Var}(B_i)=\sum_{i=1}^n\frac{1}{i}\Big(1-\frac{1}{i}\Big)=\log n+\mathcal{O}(1)$$

as $n \uparrow \infty$, we get that

$$\left(\frac{\sum_{i=1}^n B_i - \log n}{\sqrt{\log n}}, \frac{\sum_{i=1}^n B'_i - \log n}{\sqrt{\log n}}\right) \Rightarrow (\Omega_1, \Omega_2),$$

where Ω_1 and Ω_2 are two independent standard Gaussian random variables. Now,

$$\frac{|\xi(n+1)| - \log n}{\sqrt{\log n}} = \frac{\sum_{i=1}^n B'_i + \sum_{i=1}^{I_n}(B_i - B'_i) - \log n}{\sqrt{\log n}} \Rightarrow \Omega_2.$$

This concludes the proof. □

We are now ready to prove Proposition 2.7 and thus Theorem 2.3:

Proof of Proposition 2.7. The idea of the proof is that $Y_1(n+1)$ is the label of the first child in T_∞ of a node $\zeta(n+1)$ taken uniformly at random among $\{0, \dots, n\}$, and $Y_2(n+1)$ is the label of the second (in case $\xi(n+1) = \zeta(n+1)$) child in T_∞ of a node $\xi(n+1)$ taken uniformly at random among $\{0, \dots, n\}$, independently from $\zeta(n+1)$. We let K_n be the height of the last common ancestor of $\zeta(n+1)$ and $\xi(n+1)$. In distribution,

$$(Y_1(n+1), Y_2(n+1)) = (W^{(1)}_{|\zeta(n+1)|+1-K_n}, W^{(2)}_{|\xi(n+1)|+1-K_n}),$$

where $W^{(1)}$ and $W^{(2)}$ are two independent Markov chains of transition probabilities R started at the same value W_{K_n}. By Assumption on W, we thus have that, conditionally on K_n and W_{K_n},

$$\left(\frac{W^{(1)}_n - a(n)}{b(n)}, \frac{W^{(2)}_n - a(n)}{b(n)}\right) \Rightarrow (\Gamma_1, \Gamma_2),$$

in distribution as $n \uparrow \infty$, where $\Gamma_1, \Gamma_2 \sim \nu$ are independent. Because the distribution of the limit does not depend on K_n and W_{K_n}, this implies that (without conditioning)

$$\left(\frac{W^{(1)}_n - a(n)}{b(n)}, \frac{W^{(2)}_n - a(n)}{b(n)}\right) \Rightarrow (\Gamma_1, \Gamma_2). \tag{2.10}$$

Furthermore, Lemma 2.8,

$$\left(\frac{|\zeta(n+1)| - \log n}{\sqrt{\log n}}, \frac{|\xi(n+1)| - \log n}{\sqrt{\log n}}, K_n\right) \Rightarrow (\Omega_1, \Omega_2, K).$$

This implies

$$\left(\frac{|\zeta(n+1)| + 1 - K_n - \log n}{\sqrt{\log n}}, \frac{|\xi(n+1)| + 1 - K_n - \log n}{\sqrt{\log n}}\right) \Rightarrow (\Omega_1, \Omega_2). \tag{2.11}$$

Because the random recursive tree is independent from $W^{(1)}$ and $W^{(2)}$, Equations (2.10) and (2.11) hold jointly with (Γ_1, Γ_2) independent of (Ω_1, Ω_2). By Kolmogorov's representation theorem, there exists a probability space (which we call Kolmogorov's space in what follows) on which both (2.10) and (2.11) hold almost surely. The same calculation as in (2.5) gives

$$\left(\frac{Y_1(n+1) - a(\log n)}{b(\log n)}, \frac{Y_2(n+1) - a(\log n)}{b(\log n)}\right) \to \big(f(\Omega_1) + g(\Omega_1)\Gamma_1, f(\Omega_2) + g(\Omega_2)\Gamma_2\big),$$

almost surely as $n \uparrow \infty$ on Kolmogorov's probability space, and thus in distribution on the original probability space, as desired. □

2.4. **The non-balanced "irreducible" case.** This section is based on Mailler and Villemonais [MV20]. In this paper, the authors proved almost sure convergence of a large class of MVPPs; their result, however is neither stronger nor weaker than those of [MM17] discussed in Section 2.3. Indeed, an improvement with respect to [MM17] is that [MV20] allows the MVPP to be unbalanced. Also, [MV20] gets almost sure convergence of the MVPP, as opposed to convergence in probability in [MM17]. However, [MV20] only allows the Markov chain of transition probabilities R to be $(0, 1, \nu)$-ergodic (see Definition 2.3.1), so many MVPPs that fall into the framework of [MM17] are not covered by [MV20].

Also, the main difference between [MV20] and [MM17] is their methods of proof: [MM17] used what could be called a many-to-two method from the theory of branching random walks, while [MV20] uses stochastic approximation methods. To get rid of the balance assumption, [MV20] relies on the theory of quasi-stationarity for killed Markov chains.

Before stating and proving the results of [MV20], we show that finitely-many-colour Pólya urns are stochastic approximations and give yet another proof of 1.5 using results from the literature on stochastic approximations.

2.4.1. *Pólya urns are stochastic approximations.*

Definition. A stochastic approximation on subset $\mathcal{S}$ of $\mathbb{R}^d$ is a sequence $(X_n)_{n\geqslant 0}$ of $\mathcal{S}$-valued random variables such that, for some filtration $(\mathcal{F}_n)_{n\geqslant 0}$, for all $n \geqslant 0$, almost surely,

$$X_{n+1} = X_n + \gamma_n\big(F(X_n) + \Delta M_{n+1} + \varepsilon_{n+1}\big),$$

where $(\gamma_n)_{n\geqslant 0}$ is a sequence satisfying

$$\sum_{n\geqslant 0} \gamma_n = \infty \quad \text{and} \quad \sum_{n\geqslant 0} \gamma_n^2 < \infty,$$

F is a function from $\mathcal{S}$ into itself, $(\Delta M_n)_{n\geqslant 0}$ is an $(\mathcal{F}_n)_{n\geqslant 0}$-adapted sequence of martingale increments, and $(\varepsilon_n)_{n\geqslant 0}$ is an $(\mathcal{F}_n)_{n\geqslant 0}$-adapted sequence of random variables satisfying $\lim_{n\uparrow\infty} \varepsilon_n = 0$ almost surely.

In the literature, one can find many different definition for stochastic approximations, and also, many different theorems that give almost sure convergence of a stochastic approximation, under different sets of assumptions. Classical references on stochastic references include the book of [Duf97], the lecture notes of Benaïm [Ben99], and the survey paper of Pemantle [Pem07]. The heuristic general idea is that, if F is regular enough, then $(X_n)_{n\geqslant 0}$ follows the flow of the differential equation $\dot{y} = F(y)$ (this is made rigorous by the concept of "pseudo-asymptotic trajectories" of Benaïm [Ben99]), and this implies that $(X_n)_{n\geqslant 0}$ converges almost surely to the set of "stable" zeros of F.

Lemma 2.9. *Let $(U(n))_{n\geqslant 0}$ be the d-colour Pólya urn of initial composition $\alpha = (\alpha_1, \dots, \alpha_d)$ and replacement matrix $\mathfrak{r} = (\mathfrak{r}_{ij})_{1\leqslant i,j\leqslant d}$. For all $n \geqslant 0$, set $T(n) = \|U(n)\|_1$ and $\hat{U}(n) = U(n)/T(n)$. Then, for all $n \geqslant 0$,*

$$\hat{U}(n+1) = \hat{U}(n) + \gamma_n(F(\hat{U}(n)) + \Delta M_{n+1}),$$

where $\gamma_n = \frac{1}{T(n+1)}$ and F is a function from $\mathcal{S}$ into itself where

$$\mathcal{S} = \{(x_1, \dots, x_d) \in [0,1]^d \colon \sum_{i=1}^d x_i = 1\},$$

and $F(x) = \sum_{i=1}^d x_i(\mathfrak{r}_i - \|\mathfrak{r}_i\|_1 x)$ for all $x \in \mathcal{S}$.

Proof. Let $(\mathcal{F}_n)_{n\geqslant 0}$ be the natural filtration of $(U(n))_{n\geqslant 0}$. By definition, for all $n \geqslant 0$, $U(n+1) = U(n) + \mathfrak{r}_{Y(n+1)}$, where, for all $1 \leqslant i \leqslant d$

$$\mathbb{P}(Y(n+1) = i|\mathcal{F}_n) = \hat{U}_i(n).$$

Thus, for all $n \geqslant 0$,

$$\begin{aligned}\hat{U}(n+1) &= \frac{U(n) + \mathfrak{r}_{Y(n+1)}}{T(n+1)} = \hat{U}(n)\cdot\frac{T(n)}{T(n+1)} + \frac{\mathfrak{r}_{Y(n+1)}}{T(n+1)}\\ &= \hat{U}(n)\cdot\Big(1 - \frac{\|\mathfrak{r}_{Y(n+1)}\|_1}{T(n+1)} + \frac{\mathfrak{r}_{Y(n+1)}}{T(n+1)} = \hat{U}(n) + \frac{\mathfrak{r}_{Y(n+1)} - \|\mathfrak{r}_{Y(n+1)}\|_1\hat{U}(n)}{T(n+1)}.\end{aligned}$$

We set

$$\Delta M_{n+1} = \mathfrak{r}_{Y(n+1)} - \|\mathfrak{r}_{Y(n+1)}\|_1\hat{U}(n) - \mathbb{E}[\mathfrak{r}_{Y(n+1)} - \|\mathfrak{r}_{Y(n+1)}\|_1\hat{U}(n)|\mathcal{F}_n].$$

Because

$$\mathbb{E}[\mathfrak{r}_{Y(n+1)} - \|\mathfrak{r}_{Y(n+1)}\|_1\hat{U}(n)|\mathcal{F}_n] = \sum_{i=1}^d \hat{U}_i(n)(\mathfrak{r}_i - \|\mathfrak{r}_i\|_1\hat{U}(n)),$$

we get

$$\hat{U}(n+1) = \hat{U}(n) + \frac{1}{T(n+1)}\big(F(\hat{U}(n)) + \Delta M_{n+1}\big),$$

where $F : \mathbb{R}^d \to \mathbb{R}^d$ is defined by $F(x) = \sum_{i=1}^d x_i(\mathfrak{r}_i - \|\mathfrak{r}_i\|_1 x)$, as desired. □

Note that, if $\mathfrak{r}$ is balanced, i.e. for all $1 \leqslant i \leqslant d$, $\|\mathfrak{r}_i\|_1 = \mathfrak{s}$ for some $\mathfrak{s} > 0$, then $T(n) = \|\alpha\|_1 + n\mathfrak{s}$ for all $n \geqslant 0$, and $F(x) = x(R - \mathfrak{s}\mathrm{Id})$. In that case, the following result applies:

Theorem 2.10 ([Duf97, Theorem 1.4.26]). *Assume that $(X_n)_{n\geqslant 0}$ is a stochastic approximation on $\mathcal{S} \subset \mathbb{R}^d$ in the sense of Definition 2.4.1. Assume that: there exists a function $\sigma^2 : \mathcal{S} \to \mathbb{R}$ such that*

(i) F is continuous,
(ii) $\mathbb{E}[(\Delta M_{n+1} + \varepsilon_{n+1})^2|\mathcal{F}_n] \leqslant \sigma^2(X_n)$ for all $n \geqslant 0$,
(iii) there exists a constant $K > 0$ such that, for all $x \in \mathbb{R}^d$, $\|F(x)\| + \sigma^2(x) \leqslant K(1 + \|x\|^2)$,
(iv) there exists $x^* \in \mathbb{R}^d$ such that $F(x^*) = 0$ and, for all $x \in \mathcal{S}\backslash\{x^*\}$, $\langle F(x), x - x^*\rangle < 0$.

Then, almost surely as $n \uparrow \infty$, $X_n \to x^*$.

We now show how to apply Theorem 2.10 to re-prove Theorem 1.5 in the case of a balanced Pólya urn of irreducible replacement matrix $\mathfrak{r}$. In that case, by Lemma 2.9 $\hat{U}(n)$ is a stochastic approximation with $\gamma_n = 1/(\|\alpha\|_1 + n\mathfrak{s})$, $F : x \mapsto x(R - \mathfrak{s}\mathrm{Id})$,

$$\Delta M_{n+1} = \mathfrak{r}_{Y(n+1)} - \mathbb{E}[\mathfrak{r}_{Y(n+1)}|\mathcal{F}_n],$$

and $\varepsilon_{n+1} = 0$. By the triangular inequality, $\|\Delta M_{n+1}\|_1 \leqslant 2\mathfrak{s}$; we thus let $\sigma^2(x) = 2\mathfrak{s}$ for all $x \in \mathbb{R}^d$ so that Assumption (ii) of Theorem 2.10 holds. Now note that there exists $K > 0$ such that, for all $x \in \mathbb{R}^d$, $\|F(x)\| = \|x(R - \mathfrak{s}\mathrm{Id})\| \leqslant K\|x\|$ (indeed, one can take K to be equal to the spectral radius of $R - \mathfrak{s}\mathrm{Id}$). Thus, Assumption (iii) of Theorem 2.10 also holds. Because F is linear, it is also continuous. It only remains to check Assumption (iv): it holds because, by Perron-Frobenius theorem, there exists a unique left eigenvector $v \in \mathcal{S}$ associated to the eigenvalue $\mathfrak{s}$. Hence, $F(v) = 0$ and, for all $x \in \mathcal{S}\backslash\{v\}$,

$$\langle F(x), x - v\rangle = \langle x(R - \mathfrak{s}\mathrm{Id}), x - v\rangle = \langle (x - v)(R - \mathfrak{s}\mathrm{Id}), x - v\rangle < 0,$$

because all eigenvalues of $R - \mathfrak{s}\mathrm{Id}$ except v have negative real parts. Thus, Theorem 2.10 applies and gives Theorem 1.5, as expected.

2.4.2. *MVPPs are stochastic approximations.* First we show that $(\hat m_n = m_n/m_n(E))_{n\geqslant 0}$ is a stochastic approximation. Note that $(\hat m_n)_{n\geqslant 0}$ takes values in $\mathcal{P}(E)$, the set of probability measures on E. Before seeing $(\hat m_n)_{n\geqslant 0}$, we need to explain some notation: Let $\nu \in \mathcal{P}(E)$, $f : E \to \mathbb{R}$ be a continuous and bounded function, and R be a kernel on E. We let

$$\nu \cdot f = \int_E f \,\mathrm{d}\nu \quad \text{and} \quad (\nu R)\cdot f = \int_E R_x \cdot f \,\mathrm{d}\nu(x).$$

For any random probability measure μ, we let $\mathbb{E}[\mu]$ be the measure such that, for all continuous and bounded functions $f : E \to \mathbb{R}$,

$$\mathbb{E}[\mu \cdot f] = \mathbb{E}[\mu]\cdot f.$$

Lemma 2.11. *Let $(m_n)_{n\geqslant 0}$ be the stochastic approximation of initial composition m_0 and (deterministic) replacement kernel $R = (R_x)_{x\in E}$. Recall that, for all $i \geqslant 1$, $Y(i)$ is the colour of the ball picked at time i. For all $n \geqslant 0$, set $\hat m_n = m_n/m_n(E)$. Then, for all $n \geqslant 0$,*

$$\hat m_{n+1} = \hat m_n + \frac{1}{m_{n+1}(E)}(F(\hat m_n) + \Delta M_{n+1}),$$

where $F : \mathcal{P}(E) \to \mathcal{P}(E)$ is defined as $F(\mu) = \mu R - (\mu R)(E)\mu$, and ΔM_{n+1} is a martingale increment (meaning that, for all continuous functions f, $\Delta M_{n+1}\cdot f$ is a martingale increment).

Proof. By definition, for all $n \geqslant 0$, $m_{n+1} = m_n + R_{Y(n+1)}$. Thus,

$$\begin{aligned}\hat m_{n+1} = \frac{m_n + R_{Y(n+1)}}{m_{n+1}(E)} &= \Big(1 - \frac{R_{Y(n+1)}(E)}{m_{n+1}(E)}\Big)\hat m_n + \frac{R_{Y(n+1)}}{m_{n+1}(E)}\\ &= \hat m_n + \frac{1}{m_{n+1}(E)}\big(R_{Y(n+1)} - R_{Y(n+1)}(E)\hat m_n\big).\end{aligned}$$

Thus, if we let $(\mathcal{F}_n)_{n\geqslant 0}$ be the natural filtration of $(Y(i))_{i\geqslant 1}$, because

$$\mathbb{E}[R_{Y(n+1)} - R_{Y(n+1)}(E)\hat m_n|\mathcal{F}_n] = \hat m_n R - (\hat m_n R)(E)\hat m_n,$$

we have

$$\hat m_{n+1} = \hat m_n + \frac{1}{m_{n+1}(E)}\big(F(\hat m_n) + \Delta M_{n+1}\big),$$

where $F : \mathcal{P}(E) \to \mathcal{P}(E)$ is defined by $F(\mu) = \mu R - (\mu R)(E)\mu$, and where $\mathbb{E}[\Delta M_{n+1}|\mathcal{F}_n] = 0$, as claimed. □

As in Section 2.4.1, the balanced case is easier, although one solution around that is to consider $\eta_n = \frac{1}{n}\sum_{i=1}^n \delta_{Y(i)}$ instead of $\hat m_n$: it is also a stochastic approximation, and $\eta_n(E) = 1$ for all $n \geqslant 1$ (see [MV20]). For simplicity, we assume from now on that R is a probability kernel and thus that $(m_n)_{n\geqslant 0}$ is a balanced MVPP. In that case, we have

$$\hat m_{n+1} = \hat m_n + \frac{1}{m_0(E) + n}(F(\hat m_n) + \Delta M_{n+1}),$$

where $F(\mu) = \mu(R - \mathrm{Id})$.

We assume for now that, indeed, $(\hat m_n)_{n\geqslant 0}$ will eventually follow the flow of

$$\dot\mu = \mu(R - \mathrm{Id}), \tag{2.12}$$

and try to understand this flow. We let $(X_t)_{t\geqslant 0}$ is the jump Markov process that jumps at rate 1 and when it jumps, its new position is sampled according to the kernel R. If we let ν_t denote the distribution of X_t for all $t \geqslant 0$, then $(\nu_t)_{t\geqslant 0}$ is the solution of (2.12) with initial value ν_0. Thus, if we assume that $(X_t)_{t\geqslant 0}$ is ergodic, i.e. that there exists ν such that, for all ν_0, $\nu_t \to \nu$ as $t \uparrow \infty$ (for the weak topology), then this implies that the flow defined by (2.12) has one attractive, stable zero, which is ν. And thus, one expects that $\hat m_n \to \nu$ almost surely as $n \uparrow \infty$.

2.4.3. *Almost sure convergence of non-balanced MVPPs.* Before stating the main result of this section, we need the following definition:

Definition. Let $(X_t)_{t\geqslant 0}$ be a jump Markov process on some space $E \cup \{\partial\}$, absorbed at ∂. We say that $\nu \in \mathcal{P}(E)$ is a quasi-stationary distribution of X if, for some initial distribution $\pi \in \mathcal{P}(E)$,

$$\mathbb{P}_\pi(X_t \in \cdot | X_t \neq \partial) \to \nu,$$

weakly as $t \uparrow \infty$.

Theorem 2.12. *Let $(m_n)_{n\geqslant 0}$ be the MVPP of initial composition m_0 and replacement kernel R. We assume that*

(a) There exists $c > 0$ such that, for all $x \in E$, $0 < c \leqslant R_x(E) \leqslant 1$.
(b) There exists a function $V : E \to [1, \infty)$ such that, for all $L \geqslant 0$, $\{x \in E\colon V(x) \leqslant L\}$ is relatively compact, and, for all $x \in E$,

$$R_x \cdot V \leqslant \theta V(x) + K,$$

for some $\theta \in (0, c)$ and $K \geqslant 0$.
(c) The continuous-time jump process of sub-Markovian jump kernel $R - I$ admits a quasi-stationary distribution ν.

[+ technical assumptions, see [MV20]] Then, $\hat{m}_n \to \nu$ almost surely as $n \uparrow \infty$.

References

[AK68] Krishna B Athreya and Samuel Karlin. Embedding of urn schemes into continuous time Markov branching processes and related limit theorems. *The Annals of Mathematical Statistics*, 39(6):1801–1817, 1968. 6

[AN72] Krishna B Athreya and Peter E Ney. *Branching processes*. Springer, 1972. 10

[Ben99] Michel Benaïm. Dynamics of stochastic approximation algorithms. *Séminaire de probabilités de Strasbourg*, 33:1–68, 1999. 21

[BM73] David Blackwell and James B MacQueen. Ferguson distributions via Pólya urn schemes. *The Annals of Statistics*, 1(2):353–355, 1973. 11, 12, 14

[BT16] Antar Bandyopadhyay and Debleena Thacker. A new approach to Pólya urn schemes and its infinite color generalization. arXiv preprint arXiv:1606.05317, 2016. 11

[BT17] Antar Bandyopadhyay and Debleena Thacker. Pólya urn schemes with infinitely many colors. *Bernoulli*, 23(4B):3243–3267, 2017. 11

[CMP15] Brigitte Chauvin, Cécile Mailler, and Nicolas Pouyanne. Smoothing equations for large Pólya urns. *Journal of Theoretical Probability*, 28:923–957, 2015. 3

[Dob96] Robert P Dobrow. On the distribution of distances in recursive trees. *Journal of Applied Probability*, 33(3):749–757, 1996. 17

[Duf97] Marie Duflo. *Random Iterative Models*. Springer-Verlag, 1997. 21, 22

[EP23] F. Eggenberger and G. Pólya. Über die statistik verketetter vorgäge. *Zeitschrift für Angewandte Mathematik und Mechanik*, 1:279–289, 1923. 2

[Fer73] Thomas S Ferguson. A bayesian analysis of some nonparametric problems. *The Annals of Statistics*, pages 209–230, 1973. 12

[Jan04] Svante Janson. Functional limit theorems for multitype branching processes and generalized Pólya urns. *Stochastic Processes and Applications*, 110(2):177–245, 2004. 8

[Mah08] Hosam Mahmoud. *Pólya urn models*. CRC press, 2008. 1

[Mar17] A. A. Markov. Sur quelques formules limites du calcul des probabilités. (russian.). *Bulletin de l'Académie Impériale des Sciences*, 11(3):177–186, 1917. 2

[MB19] Cécile Mailler and Gerónimo Uribe Bravo. Random walks with preferential relocations and fading memory: a study through random recursive trees. *Journal of Statistical Mechanics: Theory and Experiment*, 2019(9):093206, 2019. 19

[MM17] Cécile Mailler and Jean-François Marckert. Measure-valued Pólya processes. *Electronic Journal of Probability*, 22:26, 2017. 11, 15, 18, 19, 21

[Mül] Noela Müller. Central limit theorem analogues for multicolour urn models. arXiv preprint arXiv:1604.02964. 5

[MV20] Cécile Mailler and Denis Villemonais. Stochastic approximation on non-compact measure spaces and application to measure-valued Pólya processes. *Annals of Applied Probability*, 30(5), 2020. 21, 23, 24

[Pem07] Robin Pemantle. A survey of random processes with reinforcement. *Probability Surveys*, 4:1–79, 2007. 21

[Wil91] David Williams. *Probability with martingales.* Cambridge university press, 1991. 3

Department of Mathematical Sciences, University of Bath, UK
Email address: `c.mailler@bath.ac.uk`