University of Zurich
Institute of Mathematics

Master Thesis

Rates of Convergence of Pólya Urns and Measure-Valued Pólya Processes

Author:
Davide Solfrini

Supervisor:
Prof. Jean Bertoin

January 2025

Abstract

This thesis explores the extension and generalization of the Pólya urn model, a probability model introduced in 1923 by F. Eggenberger and G. Pólya ([EP23]). It covers both the classic Pólya urn model with finitely many colours, and Measure Valued Pólya Processes, which extend the theory to infinitely many colours. The study provides key results, including proofs of limiting theorems for the urn composition over time, under various configurations of the replacement matrix R and infinite colour spaces. In particular, a version of the law of large numbers and a central limit theorem are presented for both the finitely many colours case and the MVPP case. Additionally, a brief introduction to the embedding of the urn model into continuous time, random recursive trees, branching Markov processes, stochastic approximations, and quasi-stationary distributions is provided. This thesis contributes to a deeper understanding of the Pólya urn model's mathematical structure and its potential applications, with the aid of Martingale theory and Markov chain properties. Proofs and results rely significantly on the foundational work of C. Mailler ([Mai23]).

Acknowledgements

I would like to express my sincere gratitude to Professor Bertoin for his continuous guidance and support throughout my thesis. His help, especially in clarifying concepts in probability theory, was essential. I am also grateful for his availability to answer questions and provide advice whenever I needed it.
I would also like to thank Mrs. Mailler, the author of the main paper referenced in this thesis, for her kindness and willingness to provide clarifications on her work. Her help was much appreciated and greatly contributed to the development of my research.
Finally, I want to thank my family, my girlfriend, and my friends for their constant support, encouragement, and understanding. They have been a source of strength and motivation throughout this journey. I also want to acknowledge my university colleagues for their companionship and for making this experience even more meaningful.

Contents

1 Introduction **1**
- 1.1 Historical background . . . 1
- 1.2 Brief description of the models . . . 1
  - 1.2.1 Classic Pólya urn model . . . 1
  - 1.2.2 Measure-Valued Pólya Processes . . . 2
- 1.3 Applications of Pólya urn models . . . 2
- 1.4 Main reference and objective . . . 3

2 Key results in probability theory **5**
- 2.1 Measure theory . . . 5
- 2.2 Inequalities . . . 5
- 2.3 Convergence theorems . . . 6
- 2.4 Martingale theory and Markov chains . . . 7

3 Finitely many colour Pólya urns **11**
- 3.1 Description of the model . . . 11
- 3.2 The identity matrix case . . . 12
  - 3.2.1 A law of large numbers for the identity case . . . 12
  - 3.2.2 A central limit theorem for the identity case . . . 21
- 3.3 The irreducible matrix case . . . 23
  - 3.3.1 A law of large numbers for the irreducible case . . . 23
  - 3.3.2 Embedding of an urn into continuous time . . . 24
  - 3.3.3 A law of large numbers for the continuous-time urn process . . . 25
  - 3.3.4 A central limit theorem for the irreducible case . . . 34

4 Infinitely many colour Pólya urns **57**
- 4.1 The identity kernel case . . . 58
- 4.2 The balanced irreducible case . . . 63
  - 4.2.1 Convergence in probability of irreducible MVPPs . . . 64
  - 4.2.2 Random recursive trees and branching Markov chains . . . 65
  - 4.2.3 Proof of Theorem 4.9 . . . 67
- 4.3 The non-balanced irreducible case . . . 73
  - 4.3.1 Pólya urns are stochastic approximations . . . 73
  - 4.3.2 MVPPs are stochastic approximations . . . 77

4.3.3 Quasi-stationary distributions 78
4.3.4 Almost sure convergence of non-balanced MVPPs 82
4.3.5 Examples of application of Theorem 4.33 90

5 Conclusion **93**

References **94**

Chapter 1

Introduction

1.1 Historical background

The Pólya urn model, named after the Hungarian mathematician George Pólya, was first introduced by Friedrich Eggenberger and George Pólya in their 1923 paper [EP23], and serves as a foundational concept in probability theory and statistical mechanics. Pólya's original model aimed to describe the spread of colours in an urn, where the addition of balls depended on the colours already present. This model has since been extended and generalized, contributing to various fields such as epidemiology, genetics, and computer science.
In the classic Pólya urn model, an urn contains balls of different colours. A ball is drawn at random and then replaced in the urn along with a number of additional balls of the same colour, leading to a "rich get richer" dynamic.

1.2 Brief description of the models

1.2.1 Classic Pólya urn model

In the classic Pólya urn model, we start with an urn $(U(n))_{n\geq 0}$ containing balls of different colours. Each colour i starts with $U_i(0)$ balls. At each time step n, we draw a ball from the urn, note its colour, and then put it back into the urn along with an additional ball of the same colour. This process makes it more likely to draw balls of colours that have been drawn more often before.
Mathematically, if there are d colours and the number of balls of colour i at time n is $U_i(n)$, the probability $\mathbb{P}(\xi(n+1) = i|U(n))$ of drawing a ball ξ of colour i at the next time step $n+1$ is given by:

$$\mathbb{P}(\xi(n+1) = i|U(n)) = \frac{U_i(n)}{\sum_{i=1}^{d} U_i(n)},$$

where $\sum_{i=1}^{d} U_i(n)$ is the total number of balls in the urn at time n.
These concepts will be better discussed in chapter 3.

1.2.2 Measure-Valued Pólya Processes

The idea of the Pólya urn model can be extended to situations with infinitely many colours. These are called measure-valued Pólya processes (MVPPs). In this case, instead of counting balls, we use a measure to represent the state of the urn.
In an MVPP, the colours of the balls are drawn from a potentially infinite set, and the reinforcement mechanism is described by a measure. Let m_n be the measure representing the state of the urn at time n. When a ball of colour i is drawn, the measure is updated by adding a measure $R_{Y(n+1)}$:

$$m_{n+1} = m_n + R_{Y(n+1)},$$

where $Y(n+1)$ is a random variable with a distribution dependent on the measure m_n. This update makes the total measure increase over time, and the probability of drawing a ball of a particular colour is proportional to the measure assigned to that colour.
These concepts will be better discussed in chapter 4.

1.3 Applications of Pólya urn models

We now give some examples of application of the models, along with some related references:
Finitely-Many colours

- Epidemiology: the model can describe how diseases spread through a population. Different states (like susceptible, infected, recovered) can be represented by different colours, and the transition probabilities are influenced by the current state distribution ([Mah08], [HAG17], [Pem07]).

- Genetics: it helps explain how certain traits become more or less common in a population over time due to random sampling effects ([Mah08], [BST04], [JK77]).

- Marketing: it can be used to study how products gain popularity. The more people who buy a product, the more likely it is that others will buy it too ([LK20], [Han06], [AGS23]).

Infinitely-Many colours

- Bayesian nonparametrics: the Dirichlet process, which is a cornerstone of Bayesian nonparametrics, is a kind of MVPP. It's used in scenarios where you want to model distributions without assuming a fixed number of parameters ([Pit06], [FGW+12], [Fer73]).

- Potential applications: while explicit studies on MVPPs are limited, their versatility suggests potential in many fields. They could deepen insights in ecological modelling, evolutionary biology, and social sciences. The opportunity for future research in these areas remains vast and promising.

1.4 Main reference and objective

This thesis is largely based on the work of Mrs. C. Mailler, [Mai23]. All theorems, lemmas, and other results without explicit references are taken from her insightful paper. Her contributions have been fundamental to my understanding and development of the topics discussed here.
The objective of this thesis is to prove limiting theorems for $U(n)$ as $n \to \infty$. We will investigate the convergence and fluctuations of $U(n)/n$ under two scenarios: when the replacement matrix R is a scalar multiple of the identity matrix and when R is irreducible. Additionally, we will extend our study to MVPP, where the set of colours is infinite.

Chapter 2

Key results in probability theory

We recall some important results of probability theory that will be used to prove our arguments. As a general assumption, we work on the probability space $(\Omega, \mathcal{A}, \mathbb{P})$, comprising a set Ω, a σ-field $\mathcal{A}$ of subsets of Ω, and a probability measure $\mathbb{P}$ on $(\Omega, \mathcal{A})$.

2.1 Measure theory

Theorem 2.1 (Hölder's inequality, Theorem 4.1 of [LG22]). *Let $p, q \in [1, \infty]$ with $\frac{1}{p} + \frac{1}{q} = 1$. Then, if f and g are two measurable functions from E into $\mathbb{R}$,*

$$\int |fg| \, \mathrm{d}\mu \leq \|f\|_p \|g\|_q.$$

In particular, for two random variables $X \in L^p$ and $Y \in L^q$, with $Y \in L^1$:

$$\mathbb{E}\big[|XY|\big] \leq \mathbb{E}\big[|X|^p\big]^{\frac{1}{p}} \mathbb{E}\big[|Y|^q\big]^{\frac{1}{q}}.$$

Theorem 2.2 (Jensen's inequality, Theorem 4.3 of [LG22]). *Suppose that μ is a probability measure, and let $\varphi : \mathbb{R} \to \mathbb{R}_+$ be a convex function. Then, for every $f \in L^1$,*

$$\varphi\bigg(\int f \, \mathrm{d}\mu \bigg) \leq \int \varphi \circ f \, \mathrm{d}\mu \, .$$

In particular, for a random variable $X \in L^1$:

$$\varphi\big(\mathbb{E}[X]\big) \leq \mathbb{E}\big[\varphi(X)\big].$$

2.2 Inequalities

Theorem 2.3 (Markov's inequality, Theorem 1.1, chapter 3 of [Gut13]). *Let X be a random variable. Suppose that $\mathbb{E}[|X|^p] < \infty$ for some $p > 0$, and let $c > 0$. Then,*

$$\mathbb{P}(|X| > c) \leq \frac{\mathbb{E}[|X|^p]}{c^p}.$$

Corollary 2.4 (Chebyshev's inequality, Theorem 1.4, chapter 3 of [Gut13]). *Let X be a random variable. Suppose that $\mathrm{Var}(X) < \infty$, and let $c > 0$. Then,*

$$\mathbb{P}(|X - \mathbb{E}[X]| > c) \leq \frac{\mathrm{Var}(X)}{c^2}.$$

2.3 Convergence theorems

Theorem 2.5 (Borel-Cantelli lemma, Lemma 9.11 of [LG22]). *Let $(A_n)_{n\in\mathbb{N}}$ be a sequence of events.*

i. If $\sum_{n\in\mathbb{N}} \mathbb{P}(A_n) < \infty$, then

$$\mathbb{P}(\limsup A_n) = 0.$$

ii. If $\sum_{n\in\mathbb{N}} \mathbb{P}(A_n) = \infty$ and if the events A_n, $n \in \mathbb{N}$, are independent, then

$$\mathbb{P}(\limsup A_n) = 1.$$

Theorem 2.6 (Law of Large Numbers, Theorem 10.8 of [LG22]). *Let $(X_n)_{n\geq 1}$ be a sequence of independent and identically distributed real random variables in L^1. Then,*

$$\frac{1}{n}(X_1 + \cdots + X_n) \xrightarrow{n\to\infty} \mathbb{E}[X_1] \text{ a.s.}$$

Notation. With $\xrightarrow{d}$ and $\xrightarrow{\mathbb{P}}$, we denote, respectively, convergence in distribution, and in probability.

Theorem 2.7 (Slutsky's theorem, Theorem 11.4 of [Gut13]). *Let $(X_n)_{n\geq 0}$ and $(Y_n)_{n\geq 0}$ be sequences of random variables. Suppose that*

$$X_n \xrightarrow{d} X \quad \text{and} \quad Y_n \xrightarrow{\mathbb{P}} c \quad \text{as } n \to \infty,$$

where X is a random variable and c is some constant. Then

$$\begin{aligned} X_n + Y_n &\xrightarrow{d} X + c, \\ X_n - Y_n &\xrightarrow{d} X - c, \\ X_n \cdot Y_n &\xrightarrow{d} X \cdot c, \\ \frac{X_n}{Y_n} &\xrightarrow{d} \frac{X}{c}, \quad \text{for } c \neq 0, \end{aligned}$$

as $n \to \infty$.

Theorem 2.8 (Central Limit Theorem, Theorem 10.15 of [LG22]). *Let $(X_n)_{n\geq 1}$ be a sequence of independent and identically distributed real random variables in L^2. Set $\sigma^2 = \mathrm{Var}(X_1)$ and assume that $\sigma > 0$. Then*

$$\frac{X_1 + \cdots + X_n - n\mathbb{E}[X_1]}{\sqrt{n}} \xrightarrow{d} \mathcal{N}(0, \sigma^2),$$

where $\mathcal{N}(0, \sigma^2)$ is the Gaussian distribution with mean 0 and variance σ^2.

Theorem 2.9 (Multidimensional Central Limit Theorem, Theorem 10.18 of [LG22]). *Let $(X_n)_{n\geq 1}$ be a sequence of independent and identically distributed random variables with values in $\mathbb{R}^d$, such that $\mathbb{E}[|X_1|^2] < \infty$. Then,*

$$\frac{X_1 + \cdots + X_n - n\mathbb{E}[X_1]}{\sqrt{n}} \xrightarrow{d} \mathcal{N}_d(0, K_{X_1}),$$

where $\mathcal{N}_d(0, K_{X_1})$ is the Gaussian random d-vector with mean 0 and covariance matrix $K_{X_1} = (\mathrm{Cov}(X_i, X_j))_{1\leq i,j\leq d}$

Theorem 2.10 (Skorohod's Representation Theorem, (14) of section 7.2 of [GS20]). *Let $(X_n)_{n\geq 1}$ be a sequence of random variables. If $(X_n)_{n\geq 1}$ and X, with distribution functions, respectively, F_n and F, are such that*

$$X_n \xrightarrow{d} X \text{ (or, equivalently, } F_n \to F) \text{ as } n \to \infty,$$

then there exists a probability space $(\Omega', \mathcal{A}, \mathbb{P}')$ and random variables $(Y_n)_{n\geq 1}$ and Y, mapping Ω' into $\mathbb{R}$, such that:

i. $(Y_n)_{n\geq 1}$ and Y have distribution functions, respectively, F_n and F,

ii. $Y_n \to Y$ a.s. as $n \to \infty$.

2.4 Martingale theory and Markov chains

Notation. We denote the minimum between two elements, respectively the maximum, with the symbol $\wedge$, resp. $\vee$, i.e. for two elements a, b,

$$a \wedge b \coloneqq \min(a, b) \quad \text{and} \quad a \vee b \coloneqq \max(a, b).$$

Theorem 2.11 (Optional stopping theorem, Theorem 12.12 of [LG22]). *Suppose that $(X_n)_{n\geq 0}$ is a martingale (resp. a supermartingale) and let T be a stopping time. Then $(X_{n\wedge T})_{n\geq 0}$ is also a martingale (resp. a supermartingale). In particular, if the stopping time T is bounded, one has $X_T \in L^1$, and*

$$\mathbb{E}[X_T] = \mathbb{E}[X_0] \quad (\text{resp. } \mathbb{E}[X_T] \leq \mathbb{E}[X_0]).$$

Theorem 2.12 (Martingale convergence theorem, Theorem 12.15, 12.17 and 12.27 of [LG22]). *Let $(X_n)_{n\geq 0}$ be a martingale. Then the following are equivalent:*

i. X_n converges a.s. and in L^1 to a random variable $X_\infty \in L^1$ as $n \to \infty$;

ii. The sequence $(X_n)_{n\geq 0}$ is uniformly integrable;

iii. The martingale $(X_n)_{n\geq 0}$ is closed: there exists a random variable $Z \in L^1$ such that $X_n = \mathbb{E}[Z|\mathcal{F}_n]$ for every $n \geq 0$.

Corollary 2.13 (Corollary 12.16 of [LG22]). *If $(X_n)_{n\geq 0}$ is a non-negative supermartingale, then as $n \to \infty$, $X_n \to X_\infty$ a.s. and $\mathbb{E}[X_\infty] \leq \mathbb{E}[X_0]$.*

Theorem 2.14 (Doob's L^2 martingale convergence theorem, Theorem 12.22 of [LG22]). *Let $(X_n)_{n\geq 0}$ be a martingale and $p \in (1,\infty)$. Suppose that*

$$\sup_{n\in\mathbb{N}} \mathbb{E}[|X_n|^p] < \infty,$$

then, X_n converges a.s. and in L^p to a random variable X_∞ such that

$$\mathbb{E}[|X_\infty|^p] = \sup_{n\in\mathbb{N}} \mathbb{E}[|X_n|^p].$$

Theorem 2.15 (Doob's maximal inequality in continuous time, Theorem 4.2 of [Ste01]). *Let $(X_t)_{t\geq 0}$ be a continuous non-negative submartingale and $a > 0$, then for all $p \geq 1$ we have*

$$\mathbb{P}\Big(\sup_{0\leq t\leq T} X_t > a\Big) \leq \frac{\mathbb{E}\big[X_T^p\big]}{a^p}.$$

Definition 2.16 (Definition 8.2 of [LG22]). Let X be a random variable with values in the measurable space $(E,\mathcal{E})$. We define $\mathbb{P}_X$ as the probability measure on $(E,\mathcal{E})$, defined by

$$\mathbb{P}_X(B) \coloneqq \mathbb{P}(X^{-1}(B)) = \mathbb{P}(X \in B), \quad \forall B \in \mathcal{E}.$$

Theorem 2.17 (Simple Markov Property, Theorem 13.7 of [LG22]). *Let $(\mathcal{F}_n)_{n\in\mathbb{N}}$ be the canonical filtration, i.e. $\mathcal{F}_n = \sigma(X_0, X_1, \ldots, X_n)$ for a sequence of random variable $(X_n)_{n\in\mathbb{N}}$. Let f be a non-negative measurable function on the sample space Ω and let $n \geq 0$. Then, for every $x \in E$,*

$$\mathbb{E}_x[f \circ \theta_n | \mathcal{F}_n] = \mathbb{E}_{X_n}[f],$$

where $\mathbb{E}_x$ denotes the expected values under $\mathbb{P}_x$, and, for $k \geq 0$, $\theta_k : \Omega \to \Omega$ denotes the k-shift, defined by $\theta_k((\omega_n)_{n\in\mathbb{N}}) \coloneqq (\omega_{k+n})_{n\in\mathbb{N}}$.

Theorem 2.18 (Strong Markov Property, Theorem 13.8 of [LG22]). *Under the assumptions of Theorem 2.17, let T be a stopping time of $\mathcal{F}_n$. Let f be a non-negative measurable function on the sample space Ω. Then, for every $x \in E$,*

$$\mathbb{E}_x[\mathbb{1}_{T<\infty} f \circ \theta_T | \mathcal{F}_T] = \mathbb{1}_{T<\infty} \mathbb{E}_{X_T}[f].$$

Theorem 2.19 (Chapman-Kolmogorov equations, Section 4.2 of [Ros14]). *Let X be a continuous-time Markov chain on a countable space E, with transition kernel $(P_t)_{t\geq 0}$ and generator G. Let $p_{ij} = \mathbb{P}(X(n+t) = j | X(n) = i)$, then*

$$p_{ij}(t+s) = \sum_{k\in E} p_{ik}(s) p_{kj}(t).$$

Remark 2.20. From the Chapman-Kolmogorov equations, one can derive the *forward* and *backward Kolmogorov equations* (see section 6.10 of [GS20]). Denoting P_t' as the matrix with entries $p_{ij}'(t) = \frac{\mathrm{d}}{\mathrm{d}t} p_{ij}(t)$, the forward equations are given by

$$P_t' = P_t G, \quad t \geq 0. \tag{2.1}$$

Similarly, the backwards equations are given by

$$P_t' = GP_t, \quad t \geq 0,$$

and we set $P_0 = \mathrm{Id}$.

Definition 2.21 (Quadratic variation)**.**

- Discrete-time process (pp.230 of [Kal97]): For an integer-valued stochastic process $X_t \coloneqq X(t)$, the quadratic variation of X is given by

$$[X]_t = \sum_{k \leq t} (\Delta X_k)^2,$$

  where $\Delta X_t = X_t - X_{t^-}$ is the jump at time t, and $X_{t^-} = \lim_{h \searrow 0} X(t-h)$

- Continuous-time process (Definition 2.3 of ch.I of [RY13]): For a real-valued continuous stochastic process $X_t \coloneqq X(t)$, the quadratic variation of X over the interval $[0,t]$ is defined as follows:

$$[X]_t = \lim_{n \to \infty} \sum_{i=1}^{p_n} \left(X_{t_i(n)} - X_{t_{i-1}(n)} \right)^2,$$

  where $\Delta_n = \{0 = t_0(n) < t_1(n) < \cdots < t_{p_n}(n) = t\}$ and (Δ_n) forms a refining sequence of partitions of $[0,t]$ with $|\Delta_n| \to 0$ as $n \to \infty$. This limit is defined using convergence in probability.

Theorem 2.22 (Theorem 1.3, ch.IV of [RY13])**.** *A continuous and bounded martingale M_t is of finite quadratic variation and $[M]_t$ is the unique continuous increasing adapted process vanishing at zero such that $M_t^2 - [M]_t$ is a martingale.*

A consequence of this theorem is the fact that, for a real-valued martingale X, it holds

$$\mathbb{E}[|X(t)|^2] = \mathbb{E}[[X]_t] + \mathbb{E}[|X(0)|^2]. \tag{2.2}$$

Notation (Orientation of vectors)**.** In this thesis, as well as in our primary reference [Mai23], note that we will refer to a d-dimensional vector as a horizontal vector, with its transpose being a vertical vector. For example, consider the canonical basis of $\mathbb{R}^d$, $(e_i)_{1 \leq i \leq d}$, then

$$e_1 = (1, 0, \ldots, 0) \quad \text{and} \quad e_1^T = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}.$$

Chapter 3

Finitely many colour Pólya urns

3.1 Description of the model

A d-colour Pólya urn consists of balls of up to d different colours, for $d \geq 2$. The urn evolves in discrete time steps and at each step the urn is shaken and a ball is randomly selected, with all balls being equally likely. The colour of the selected is noted and the ball is returned to the urn. If the selected ball is of colour i, for $i = 1, \ldots, d$, then $R_{i,j}$ balls of colour j are added to the urn, for all $j = 1, \ldots, d$, where $R_{i,j}$ follows a discrete distribution on a set of integers. The entries $R_{i,j}$ can be deterministic or random, positive or negative. Conventionally, this process is represented using a $d \times d$ matrix, so-called replacement matrix $R = (R_{i,j})_{1 \leq i,j \leq d}$, the rows of which are indexed by the colour of the picked balls, and the columns are indexed by the colour of the added balls.

Example 3.1. For instance, consider the 2-colour Pólya urn illustrated in Figure 3.1, where we designate colour 1 as red, and colour 2 as blue. For this model we consider the replacement matrix

$$R' = \begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix},$$

and with initial composition consisting of 1 red ball and 1 blue ball.

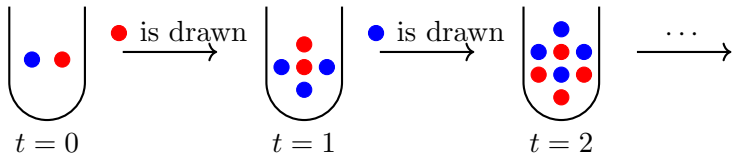

Figure 3.1: 2-colour Pólya urn model: At time $t = 0$, there is one ball per colour. A red ball is drawn, and three balls are added to the urn, according to the replacement matrix R'. The same is done at $t = 1$, where a blue ball is drawn, and the urn is filled with additional coloured balls.

Notice that a d-colour Pólya urn is a Markov process, which we will denote as $(U(n))_{n \geq 0}$, whose distribution depends on two parameters:

- the initial composition $U(0) \in \mathbb{N}^d$,
- the replacement matrix $R = (R_{i,j})_{1 \leq i,j \leq d} \in M_d(\mathbb{N})$.

One can intuitively see that it is indeed a Markov process: suppose that at time $n > 0$ we have a certain amount of balls of different colours, then the probability of picking a ball of a specific colour does not depend on the composition of $U(n_0)$ for all times $n_0 < n$, but it depends just on the present time n. In formulas, let r_n be the number of red balls at time n, and b_n the number of blue balls respectively. Let Λ_r be an event which depends on the history $(r_0, r_1, \ldots, r_{n-1})$ (or respectively Λ_b for $(b_0, \ldots, b_{n-1})$), then:

$$\mathbb{P}(r_{n+1} = x | r_n = y, \Lambda_r) = \mathbb{P}(r_{n+1} = x | r_n = y).$$

Remark 3.2. It is important to note that the replacement matrix R should not be misinterpreted as a transition probability matrix. A Pólya urn represents a time-inhomogeneous Markov chain, thus the transition probability matrix may not be constant and thus change at every time step.

We define now the distribution of $(U(n))_{n \geq 0}$ recursively: let $\xi(n+1)$ be a random variable in $\{1, \ldots, d\}$ for every $n \geq 0$ with distribution

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

where $U_i(n)$ is the i-th coordinate of the vector $U(n)$, and denotes the number of balls of colour i at time n, and $\|U(n)\|_1$ denotes the total number of balls at time n, i.e. $\|U(n)\|_1 = \sum_{i=1}^d U_i(n)$. Set $U(n+1) = U(n) + R_{\xi(n+1)}$, where R_i is the i-th line of R, for all $1 \leq i \leq d$. Reflecting what we described initially, this means that, at each time step n, we pick a ball of colour $\xi(n)$ uniformly at random, and replace it in the urn together with additional $R_{\xi(n),j}$ balls of colour j, for every $1 \leq j \leq d$.

Remark 3.3. The assumption that the entries of R and $U(0)$ should be integers is not essential for the model's description or most of its results, though it is intuitive when interpreted as balls. Instead, one could conceptualize the contents of the urn as coloured dust. The coefficients of R may also be negative, indicating that balls are removed rather than added. This approach works as long as the number of balls of each colour remains positive. If the number of balls of any colour becomes negative, we consider the urn extinct. An urn that almost surely never becomes extinct is referred to as *tenable*. One can find conditions for $U(0)$ and R to ensure tenability, see e.g. [Mah08], section 3.1. One sufficient condition for tenability is that for all $1 \leq i \leq d$, there exists κ_i such that $R_{i,i} = -\kappa_i$, and for all $1 \leq j \leq d$, $R_{i,j} > 0$ is divisible by κ_i. For simplicity, we will assume that this sufficient condition holds.

3.2 The identity matrix case

3.2.1 A law of large numbers for the identity case

Without loss of generality, we assume that for every $i \in \{1, \ldots, d\}$, $\alpha_i := U_i(0) \geq 1$.

Theorem 3.4. *Assume that $R = S \cdot Id$ and for every $i \in \{1, \dots, d\}$, $\alpha_i \geq 1$. Then*

$$\frac{U(n)}{nS} \xrightarrow{n\to\infty} V = (V_1, \dots, V_d) \quad \text{a.s.}$$

where V is Dirichlet distributed with parameters $\left(\frac{\alpha_1}{S}, \dots, \frac{\alpha_d}{S}\right)$.

Before proceeding with the proof of the theorem, we will first introduce a couple of properties of the Dirichlet distribution and its moments (see e.g. section 1, ch. 49 of [KBJ04]): Let $X = (X_1, \dots, X_d) \sim \mathrm{Dir}(\nu_1, \dots, \nu_d)$, then its density is given by

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

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

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

In particular, the Dirichlet distribution with parameters $(1, \dots, 1)$ is the uniform distribution on Σ. Moreover, for $1 \leq i \leq d$, and writing $\bar{\nu} = \sum_{k=1}^{d} \nu_k$ we have

$$\mathbb{E}[X_i] = \frac{\nu_i}{\bar{\nu}} \quad \text{and} \quad \mathrm{Var}(X_i) = \frac{\nu_i(\bar{\nu} - \nu_i)}{\bar{\nu}^2(\bar{\nu} + 1)},$$

and, in particular for $n \in \mathbb{N}^d$,

$$\mathbb{E}[X^n] = \frac{\Gamma(\bar{\nu})}{\Gamma(\bar{\nu} + |n|)} \prod_{i=1}^{d} \frac{\Gamma(\nu_i + n_i)}{\Gamma(\nu_i)},$$

where $|n| = \sum_{k=1}^{d} n_k$.

Proof of Theorem 3.4. Let e_i denote the canonical basis of $\mathbb{R}^d$ for every $1 \leq i \leq d$. For simplicity we write $\bar{\alpha} \coloneqq \sum_{i=1}^{d} \alpha_i = \|U(0)\|_1$.
Computing the conditional expectation of $U(n+1)$ knowing the previous step yields

$$\begin{aligned}
\mathbb{E}[U(n+1)|U(n)] &= \mathbb{E}[U(n) + R_{\xi_{n+1}}|U(n)] \\
&= U(n) + \mathbb{E}[R_{\xi_{n+1}}|U(n)] \\
&= U(n) + \mathbb{E}[Se_{\xi_{n+1}}|U(n)] \\
&= U(n) + S\Big(\sum_{i=1}^{d} e_i \mathbb{E}[\mathbb{1}_{\xi_{n+1}=i}|U(n)]\Big) \\
&= U(n) + S\Big(\sum_{i=1}^{d} \frac{U_i(n)}{\|U(n)\|_1} e_i\Big) \\
&= U(n)\Big(1 + \frac{S}{\|U(n)\|_1}\Big).
\end{aligned}$$

Hence, $M_n \coloneqq \frac{U(n)}{\|U(n)\|_1}$ is a $[0,1]^d$-valued martingale with mean $(\frac{\alpha_1}{\bar{\alpha}}, \dots, \frac{\alpha_d}{\bar{\alpha}})$ (Indeed, by the martingale property: $\mathbb{E}[M_n] = \mathbb{E}[M_0] = (\frac{\alpha_1}{\bar{\alpha}}, \dots, \frac{\alpha_d}{\bar{\alpha}})$). Moreover, since it is non-negative, by the Martingale Convergence Theorem corollary (Corollary 2.13), M_n converges almost surely to a random variable V. By definition, for every $n \geq 0$ we have

$$\|U(n)\|_1 = \|U(n-1) + R_{\xi_n}\|_1 = \bar{\alpha} + \sum_{i=1}^{n} R_{\xi_i} = \bar{\alpha} + nS.$$

Therefore,

$$\frac{U(n)}{nS} = \frac{U(n)}{\|U(n)\|_1} \cdot \frac{\|U(n)\|_1}{nS} = \underbrace{\frac{U(n)}{\|U(n)\|_1}}_{\longrightarrow V} \cdot \underbrace{\frac{\bar{\alpha} + nS}{nS}}_{\longrightarrow 1} \longrightarrow V \text{ a.s.}$$

Now, we want to show that V is indeed Dirichlet distributed. Let $(\mathcal{F}_n)_{n\geq 0}$ denote the natural filtration of the process $(U(n))_{n\geq 0}$ and notice that, if f is any function defined on $\mathbb{R}^d$, we have

$$\begin{aligned}
\mathbb{E}[f(U(n+1))|\mathcal{F}_n] &= \mathbb{E}[f(U(n) + Se_{\xi_{n+1}})|\mathcal{F}_n] \\
&= \mathbb{E}[f(U(n) + Se_i)\mathbb{1}_{\xi_{n+1}=i}|\mathcal{F}_n] \\
&= \sum_{i=1}^{d} f(U(n) + Se_i)\frac{U_i(n)}{\bar{\alpha} + nS}.
\end{aligned}$$

Since $\frac{1}{\bar{\alpha}+nS}\sum_{i=1}^{d} U_i(n) = \frac{\|U(n)\|_1}{\bar{\alpha}+nS} = 1$, we can write

$$\begin{aligned}
\mathbb{E}[f(U(n+1))|\mathcal{F}_n] &= \frac{1}{\bar{\alpha} + nS}\sum_{i=1}^{d} U_i(n) f(U(n) + Se_i) + f(U(n))\Big(1 - \frac{1}{\bar{\alpha} + nS}\sum_{i=1}^{d} U_i(n)\Big) \\
&= f(U(n)) + \frac{1}{\bar{\alpha} + nS}\sum_{i=1}^{d} U_i(n)[f(U(n) + Se_i) - f(U(n))] \\
&= f(U(n)) + \frac{1}{\bar{\alpha} + nS}\Phi(f)(U(n)) \\
&= \Big(\mathrm{Id} + \frac{\Phi}{\bar{\alpha} + nS}\Big)(f)(U(n)).
\end{aligned}$$

where $\Phi(f)(v) \coloneqq \sum_{i=1}^{d} v_i[f(v + Se_i) - f(v)]$ and $v = \sum_{i=1}^{d} v_i e_i$.
We introduce, for $p = (p_1, \dots, p_d) \in \mathbb{N}^d$ and $|p| = \sum_{i=1}^{d} p_i$, the function Γ_p defined on $\mathbb{R}^d$ as

$$\Gamma_p(v) \coloneqq \prod_{i=1}^{d} \frac{\Gamma(\frac{v_i}{S} + p_i)}{\Gamma(\frac{v_i}{S})}.$$

Claim. *Γ_p is an eigenfunction of the operator Φ with associated eigenvalue $|p|S$.*

Proof. We do a straightforward computation for $v \in \mathbb{R}^d$ using the Gamma function property $\Gamma(z+1) = z\Gamma(z)$, and denoting by $(e_i)_k$ the k-th entry of the vector e_i, so that for the canonical basis $(e_i)_{1\le i\le d}$ we have $(e_i)_k = \begin{cases} 1 & \text{if } i = k \\ 0 & \text{if } i \neq k \end{cases}$.

$$\begin{aligned}
\Phi(\Gamma_p)(v) &= \sum_{i=i}^{d} v_i[\Gamma_p(v + Se_i) - \Gamma_p(v)] \\
&= \sum_{i=1}^{d} v_i \left[\prod_{k=1}^{d} \frac{\Gamma(\frac{v_k}{S} + (e_i)_k + p_k)}{\Gamma(\frac{v_k}{S} + (e_i)_k)} - \prod_{k=1}^{d} \frac{\Gamma(\frac{v_k}{S} + p_k)}{\Gamma(\frac{v_k}{S})} \right] \\
&= \sum_{i=1}^{d} v_i \left[\prod_{\substack{k=1 \\ k\neq i}}^{d} \frac{\Gamma(\frac{v_k}{S} + p_k)}{\Gamma(\frac{v_k}{S})} \left(\frac{\Gamma(\frac{v_i}{S} + p_i + 1)}{\Gamma(\frac{v_i}{S} + 1)} - \frac{\Gamma(\frac{v_i}{S} + p_i)}{\Gamma(\frac{v_i}{S})} \right) \right] \\
&= \sum_{i=1}^{d} v_i \left[\prod_{\substack{k=1 \\ k\neq i}}^{d} \frac{\Gamma(\frac{v_k}{S} + p_k)}{\Gamma(\frac{v_k}{S})} \left(\frac{(\frac{v_i}{S} + p_i)\Gamma(\frac{v_i}{S} + p_i) - \frac{v_i}{S}\Gamma(\frac{v_i}{S} + p_i)}{\frac{v_i}{S}\Gamma(\frac{v_i}{S})} \right) \right] \\
&= \sum_{i=1}^{d} v_i \left[\prod_{\substack{k=1 \\ k\neq i}}^{d} \frac{\Gamma(\frac{v_k}{S} + p_k)}{\Gamma(\frac{v_k}{S})} \left(\frac{S \cdot p_i}{v_i} \cdot \frac{\Gamma(\frac{v_i}{S} + p_i)}{\Gamma(\frac{v_i}{S})} \right) \right] \\
&= S \sum_{i=1}^{d} p_i \cdot \prod_{k=1}^{d} \frac{\Gamma(\frac{v_k}{S} + p_k)}{\Gamma(\frac{v_k}{S})} \\
&= |p| S \cdot \Gamma_p(v).
\end{aligned}$$

■

Thus, by the claim, we can write

$$\begin{aligned}
\mathbb{E}[\Gamma_p(U(n+1))|\mathcal{F}_n] &= \Gamma_p(U(n)) + \frac{1}{\bar{\alpha} + nS}\Phi(\Gamma_p)(U(n)) \\
&= \Big(\frac{|p|S}{\bar{\alpha} + nS} + 1\Big)\Gamma_p(U(n)).
\end{aligned}$$

By the law of total expectation, taking expectation on both sides yields

$$\mathbb{E}[\Gamma_p(U(n+1))] = \Big(\frac{|p|S}{\bar{\alpha} + nS} + 1\Big)\mathbb{E}[\Gamma_p(U(n))].$$

Iterating this procedure, we get

$$\begin{aligned}
\mathbb{E}[\Gamma_p(U(n+1))] &= \Big(\frac{|p|S}{\bar{\alpha} + nS} + 1\Big)\Big(\frac{|p|S}{\bar{\alpha} + (n-1)S} + 1\Big)\mathbb{E}[\Gamma_p(U(n-1))] \\
&= \cdots = \mathbb{E}[\Gamma_p(U(0))] \cdot \prod_{i=0}^{n} \Big(\frac{|p|S}{\bar{\alpha} + iS} + 1\Big) \qquad (3.1) \\
&= \Gamma_p(U(0)) \cdot \prod_{i=0}^{n} \Big(\frac{|p|S}{\bar{\alpha} + iS} + 1\Big),
\end{aligned}$$

since $\Gamma_p(U(0)) = \prod_{i=0}^{d} \frac{\Gamma(\frac{\alpha_i}{S}+p_i)}{\Gamma(\frac{\alpha_i}{S})}$ is constant.
Now we prove by induction that it holds

$$\mathbb{E}[\Gamma_p(U(n))] = \frac{\Gamma(\frac{\bar{\alpha}}{S}+n+|p|)}{\Gamma(\frac{\bar{\alpha}}{S}+n)} \cdot \frac{\Gamma(\frac{\bar{\alpha}}{S})}{\Gamma(\frac{\bar{\alpha}}{S}+|p|)} \cdot \Gamma_p(U(0)) \tag{3.2}$$

for every $n \in \mathbb{N}$ (this will be our *Induction Hypothesis* (IH)).
For $n = 0$ it is straightforward. For $n = 1$, using the property $\Gamma(z+1) = z\Gamma(z)$, we have

$$\begin{aligned}
\mathbb{E}[\Gamma_p(U(1))] &= \frac{\Gamma(\frac{\bar{\alpha}}{S}+1+|p|)}{\Gamma(\frac{\bar{\alpha}}{S}+1)} \cdot \frac{\Gamma(\frac{\bar{\alpha}}{S})}{\Gamma(\frac{\bar{\alpha}}{S}+|p|)} \cdot \Gamma_p(U(0)) \\
&= \frac{(\frac{\bar{\alpha}}{S}+|p|)\Gamma(\frac{\bar{\alpha}}{S}+|p|)}{\frac{\bar{\alpha}}{S}\Gamma(\frac{\bar{\alpha}}{S})} \cdot \frac{\Gamma(\frac{\bar{\alpha}}{S})}{\Gamma(\frac{\bar{\alpha}}{S}+|p|)} \cdot \Gamma_p(U(0)) \\
&= \Big(1 + \frac{|p|S}{\bar{\alpha}}\Big)\Gamma_p(U(0)),
\end{aligned}$$

which corresponds with the result given by equation (3.1).
The inductive step $n \mapsto n+1$ gives

$$\begin{aligned}
\mathbb{E}[\Gamma_p(U(n+1))] &= \Big(\frac{|p|S}{\bar{\alpha}+nS}+1\Big)\mathbb{E}[\Gamma_p(U(n))] \\
&\overset{\text{(IH)}}{=} \Big(\frac{|p|S}{\bar{\alpha}+nS}+1\Big) \cdot \frac{\Gamma(\frac{\bar{\alpha}}{S}+n+|p|)}{\Gamma(\frac{\bar{\alpha}}{S}+n)} \cdot \frac{\Gamma(\frac{\bar{\alpha}}{S})}{\Gamma(\frac{\bar{\alpha}}{S}+|p|)} \cdot \Gamma_p(U(0)) \\
&= \frac{(\frac{\bar{\alpha}}{S}+n+|p|)}{(\frac{\bar{\alpha}}{S}+n)} \cdot \frac{\Gamma(\frac{\bar{\alpha}}{S}+n+|p|)}{\Gamma(\frac{\bar{\alpha}}{S}+n)} \cdot \frac{\Gamma(\frac{\bar{\alpha}}{S})}{\Gamma(\frac{\bar{\alpha}}{S}+|p|)} \cdot \Gamma_p(U(0)) \\
&= \frac{\Gamma(\frac{\bar{\alpha}}{S}+(n+1)+|p|)}{\Gamma(\frac{\bar{\alpha}}{S}+(n+1))} \cdot \frac{\Gamma(\frac{\bar{\alpha}}{S})}{\Gamma(\frac{\bar{\alpha}}{S}+|p|)} \cdot \Gamma_p(U(0)).
\end{aligned}$$

Therefore, formula (3.2) holds for all $n \in \mathbb{N}$. Consequently, recalling *Stirling's approximation* for the Gamma function $\Gamma(z) = \sqrt{\frac{2\pi}{z}}\big(\frac{z}{e}\big)^z\Big(1+O(\frac{1}{z})\Big)$ for any complex z large enough in absolute value, as n tends to infinity, we get

$$\frac{\Gamma(n+\frac{\bar{\alpha}}{S}+|p|)}{\Gamma(n+\frac{\bar{\alpha}}{S})} = n^{|p|}\Big(1+O\Big(\frac{1}{n}\Big)\Big),$$

by the relation, also discussed by Rowe [Row31],

$$\frac{\Gamma(x+a)}{\Gamma(x+b)} \approx x^{a-b}.$$

Therefore as $n \to \infty$, by equation (3.2), it holds

$$\mathbb{E}[\Gamma_p(U(n))] = n^{|p|} \cdot \frac{\Gamma(\frac{\bar{\alpha}}{S})}{\Gamma(\frac{\bar{\alpha}}{S}+|p|)} \cdot \Gamma_p(U(0))\Big(1+O\Big(\frac{1}{n}\Big)\Big).$$

On the other hand, we can write monomials X^p in terms of Γ_p polynomials. For this, let $X = (X_1, \dots, X_d)$ and notice that by repeatedly applying $\Gamma(z+1) = z\Gamma(z)$, we have

$$\Gamma_p(X) = \prod_{i=1}^{d} \left(\prod_{k=0}^{p_i - 1} \left(\frac{X_i}{S} + k \right) \right).$$

So we can use *Stirling numbers of the second kind* $S(p, i)$, where

$$S(p,i) := \frac{1}{i!} \sum_{k=0}^{i} (-1)^{i-k} \binom{i}{k} k^p,$$

in order to expand real polynomials $X^p = X_1^{p_1} \cdots X_d^{p_d}$ in the basis $(\Gamma_p)_{p \in \mathbb{N}^d}$. Stirling numbers of the second kind are useful to express a variable x to the n-th power in terms of $S(n, i)$, see e.g. [Gra94], section 6.1. It is enough to have for a single variable $x \in X = (X_1, \dots, X_d)$ and $n \in \mathbb{N}$ the formula

$$x^n = \sum_{i=0}^{n} (-1)^{n-i} S(n,i) x(x+1) \cdots (x+i-1),$$

such that, when applied to $\left(\frac{x}{S}\right)^n$ we get

$$x^n = S^n \sum_{i=0}^{n} \underbrace{(-1)^{n-i} S(n,i)}_{\in \mathbb{Q}} \frac{x}{S} \left(\frac{x}{S} + 1\right) \cdots \left(\frac{x}{S} + i - 1\right),$$

and therefore for $X = (X_1, \dots, X_d) \in \mathbb{R}^d$ and $p = (p_1, \dots, p_n \in \mathbb{N}^d)$:

$$X^p = S^{|p|} \Gamma_p(X) + \sum_{\substack{i \in \mathbb{N}^d \\ |i| \leq |p| - 1}} a_{p,i} \Gamma_i(X),$$

where $a_{p,i}$ are rational numbers. Then,

$$\mathbb{E}[U(n)^p] = S^{|p|} \mathbb{E}[\Gamma_p(U(n))] + \sum_{\substack{i \in \mathbb{N}^d \\ |i| \leq |p| - 1}} a_{p,i} \mathbb{E}[\Gamma_i(U(n))],$$

where the rightmost term on the right-hand side goes to $O(n^{p-1})$ as n goes to infinity, so that, dividing by $(\bar{\alpha} + nS)^p$, gives

$$\mathbb{E}\left[\left(\frac{U(n)}{\bar{\alpha} + nS} \right)^p \right] = \frac{\Gamma(\frac{\bar{\alpha}}{S})}{\Gamma(\frac{\bar{\alpha}}{S} + |p|)} \cdot \Gamma_p(U(0)) \left(1 + O\left(\frac{1}{n} \right) \right),$$

which implies that, for any $p \in \mathbb{N}^d$,

$$\mathbb{E}[V^p] = \frac{\Gamma(\frac{\bar{\alpha}}{S})}{\Gamma(\frac{\bar{\alpha}}{S} + |p|)} \cdot \Gamma_p(U(0)) = \frac{\Gamma(\frac{\bar{\alpha}}{S})}{\Gamma(\frac{\bar{\alpha}}{S} + |p|)} \prod_{i=1}^{d} \frac{\Gamma(\frac{\alpha_i}{S} + p_i)}{\Gamma(\frac{\alpha_i}{S})}.$$

By the work of Chauvin, Mailler, Pouyanne [CMP13], we know that a Dirichlet distribution is determined by its moments, so we conclude that $V \sim \mathrm{Dir}(\frac{\alpha_1}{S}, \dots, \frac{\alpha_d}{S})$. □

Remark 3.5. Notice that the last derivation in the proof of Theorem 3.4 proves the convergence of the martingale $M_n = \frac{U(n)}{\|U(n)\|_1}$ in L^t for any $t \geq 1$.

Corollary 3.6. *Assume that $d = 2$ and $R = \mathrm{Id}$.*

i. If $\alpha_1 = \alpha_2 = 1$, then

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

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

ii. If $\alpha_1, \alpha_2 \geq 1$, then

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

almost surely, where $B \sim \mathrm{Beta}(\alpha_1, \alpha_2)$.

Proof. The convergence to a random vector follows directly from Theorem 3.4, so we dedicate ourselves in proving the distributions of X and B.
So, we consider X first, and we aim to compute $\mathbb{P}(U_1(n) = k)$. To avoid any confusion within our reasoning, we say that colour 1 of the ball in the urn is red and that colour 2 is blue. For $1 \leq k \leq n \in \mathbb{N}$, in order to have k red balls at time n, since at time 0 we have 1 red ball and 1 blue ball, we would have to pick a red ball $k-1$ times, and therefore we should have picked a blue ball for the remaining $n-(k-1)$ times. The probability to do this is

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

If the draws are done in a different order, only the order of the numerators changes, but the counting stays the same overall, independently of the drawing order of the $k-1$ red balls and the $n-k+1$ blue balls. There are $\binom{n}{k-1}$ such orders, so we get

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

That is, $U_1(n)$ is uniform on $\{1,\dots,n+1\}$ for every $n \geq 1$, which implies that $X \sim \mathrm{Unif}(0,1)$.
For part (ii) (also see Theorem 2.3 of [Mah08]) we consider the distribution of B. Recall that the density of the Beta distribution with parameter (α,β) is given by $\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}x^{\alpha-1}(1-x)^{\beta-1}$.
Notice that if $\alpha_1 = \alpha_2 = 1$, the statement holds since the Beta distribution with parameters $(1,1)$ corresponds to the uniform distribution. Let $\alpha_1, \alpha_2 \geq 1$, we argue analogously as in the previous part to compute $\mathbb{P}(U_i(n) = k)$ and we have

$$\begin{aligned}\mathbb{P}(U_1(n) = k) = \binom{n}{k}&\frac{\alpha_1}{\alpha_1+\alpha_2}\cdot\frac{\alpha_1+1}{\alpha_1+\alpha_2+1}\cdots\frac{\alpha_1+k-1}{\alpha_1+\alpha_2+k-1}\\ &\cdot\frac{\alpha_2}{\alpha_1+\alpha_2+k}\cdots\frac{\alpha_2+n-k-1}{\alpha_1+\alpha_2+n-1}.\end{aligned}$$

We can write the latter equation using Gamma function as follows:

$$\begin{aligned}\mathbb{P}(U_1(n)=k) &= \frac{\Gamma(n+1)}{\Gamma(n-k+1)\Gamma(k+1)}\cdot\frac{\Gamma(\alpha_1+k)}{\Gamma(\alpha_1)}\cdot\frac{\Gamma(\alpha_1+\alpha_2)}{\Gamma(\alpha_1+\alpha_2+k)}\\ &\quad\cdot\frac{\Gamma(\alpha_2+n-k)}{\Gamma(\alpha_2)}\cdot\frac{\Gamma(\alpha_1+\alpha_2+k)}{\Gamma(\alpha_1+\alpha_2+n)}\\ &= \frac{\Gamma(\alpha_1+\alpha_2)}{\Gamma(\alpha_1)\Gamma(\alpha_2)}\cdot\frac{\Gamma(k+\alpha_1)}{\Gamma(k+1)}\cdot\frac{\Gamma(n-k+\alpha_2)}{\Gamma(n-k+1)}\cdot\frac{\Gamma(n+1)}{\Gamma(n+\alpha_1+\alpha_2)}\\ &\approx \frac{\Gamma(\alpha_1+\alpha_2)}{\Gamma(\alpha_1)\Gamma(\alpha_2)}\cdot k^{\alpha_1-1}(n-k)^{\alpha_2-1}n^{1-\alpha_1-\alpha_2},\end{aligned}$$

where in the last approximation we used Stirling's approximation.
So, as k goes to infinity, but $k \le nx$ for some $x \in (0,1)$,

$$\mathbb{P}(U_1(n)=k) = \frac{\Gamma(\alpha_1+\alpha_2)}{\Gamma(\alpha_1)\Gamma(\alpha_2)}\cdot k^{\alpha_1-1}(n-k)^{\alpha_2-1}n^{1-\alpha_1-\alpha_2}. \tag{3.3}$$

Now notice that

$$\mathbb{P}\bigg(\frac{U_1(n)}{n}\le x\bigg) = \mathbb{P}\bigg(\frac{U_1(n)}{n}=0\bigg)+\mathbb{P}\bigg(\frac{U_1(n)}{n}=\frac{1}{n}\bigg)+\cdots+\mathbb{P}\bigg(\frac{U_1(n)}{n}=\frac{\lfloor nx\rfloor}{n}\bigg).$$

Also, since $\frac{1}{n}\to 0$ as $n\to\infty$:

$$\int_0^x \mathbb{P}\bigg(\frac{U_1(n)}{n}=u\bigg)\,\mathrm{d}u = \lim_{n\to\infty}\frac{1}{n}\bigg[\mathbb{P}\bigg(\frac{U_1(n)}{n}=0\bigg)+\cdots+\mathbb{P}\bigg(\frac{U_1(n)}{n}=\frac{\lfloor nx\rfloor}{n}\bigg)\bigg],$$

so that we have

$$\mathbb{P}\bigg(\frac{U_1(n)}{n}\le x\bigg) = n\int_0^x \mathbb{P}\bigg(\frac{U_1(n)}{n}=u\bigg)\,\mathrm{d}u = n\int_0^x \mathbb{P}\big(U_1(n)=nu\big)\,\mathrm{d}u\,,$$

and since, as n tends to infinity, $nu\to\infty$, we can use equation (3.3), to get

$$\begin{aligned}\mathbb{P}\bigg(\frac{U_1(n)}{n}\le x\bigg) &= n\int_0^x \frac{\Gamma(\alpha_1+\alpha_2)}{\Gamma(\alpha_1)\Gamma(\alpha_2)}\cdot n^{\alpha_1-1}u^{\alpha_1-1}(n-nu)^{\alpha_2-1}n^{1-\alpha_1-\alpha_2}\,\mathrm{d}u\\ &= \int_0^x \frac{\Gamma(\alpha_1+\alpha_2)}{\Gamma(\alpha_1)\Gamma(\alpha_2)}\cdot u^{\alpha_1-1}(1-u)^{\alpha_2-1}\,\mathrm{d}u\,,\end{aligned}$$

which is the cumulative distribution function of a Beta distributed random variable with parameters (α_1,α_2), and since cumulative distribution functions characterize the law of a random variable, we are done. □

Theorem 3.7. *Fix $\alpha_1,\dots,\alpha_d \ge 1$ and $S \ge 1$. Define the d-dimensional process $(Z(n))_{n\ge 0}$ as follows: let V be a random variable with $V\sim \mathrm{Dir}(\frac{\alpha_1}{S},\dots,\frac{\alpha_d}{S})$, let $(X_m)_{m\ge 1}$ be a sequence of random variables such that for every $k\ge 1$, for every $1\le i\le d$,*

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

Let $Z_i(0) = \alpha_i \geq 1$ for every $1 \leq i \leq d$, and for every $n \geq 1$,

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

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

Proof. We want to compute the probability distribution of $U_i(n)$ and $Z_i(n)$ for every $1 \leq i \leq d$ to show that they are equal, so that $(Z(n))_{n\geq 0}$ is indeed distributed as the Pólya urn of initial composition $(\alpha_1, \ldots, \alpha_d)$ and replacement matrix $S \cdot \mathrm{Id}$. So, the computations yield:

- Notice that, by the law of total expectation, for any integer k and random variables X, V, it holds

$$\mathbb{P}(X = k) = \mathbb{E}[\mathbb{1}_{X=k}] = \mathbb{E}[\mathbb{E}[\mathbb{1}_{X=k}|V]] = \mathbb{E}[\mathbb{P}(X = k|V)]. \tag{3.4}$$

  Thus,

$$\mathbb{P}(Z_i(n) = \alpha_i + Sk_i \ (\forall 1 \geq i \geq d)) \overset{(3.4)}{=} \mathbb{E}[\mathbb{P}(Z_i(n) = \alpha_i + Sk_i \ (\forall 1 \geq i \geq d)|V)],$$

  where, by the multinomial distribution,

$$\mathbb{P}(Z_i(n) = \alpha_i + Sk_i \ (\forall 1 \leq i \leq d)|V) = \binom{n}{k_1, \ldots, k_d} \prod_{i=1}^{d} V_i^{k_i} \mathbb{1}_{k_1+\cdots+k_d=n}.$$

  Thus,

$$\begin{aligned}
\mathbb{P}(Z_i(n) = \alpha_i + Sk_i \ (\forall 1 \geq i \geq d)) &= \mathbb{E}\left[\frac{n!}{k_1! \cdots k_d!} \prod_{i=1}^{d} V_i^{k_i} \mathbb{1}_{k_1+\cdots+k_d=n}\right] \\
&= \frac{n!}{k_1! \cdots k_d!} \mathbb{E}\left[\prod_{i=1}^{d} V_i^{k_i}\right] \mathbb{1}_{k_1+\cdots+k_d=n} \\
&= \frac{n!}{k_1! \cdots k_d!} \cdot \frac{\Gamma(\frac{\alpha_1}{S} + \cdots + \frac{\alpha_d}{S})}{\Gamma(\frac{\alpha_1}{S} + \cdots + \frac{\alpha_d}{S} + k_1 + \cdots + k_d)} \cdot \prod_{i=1}^{d} \frac{\Gamma(\frac{\alpha_i}{S} + k_i)}{\Gamma(\frac{\alpha_i}{S})} \mathbb{1}_{k_1+\cdots+k_d=n} \\
&= \frac{n!}{k_1! \cdots k_d!} \cdot \frac{\Gamma(\frac{\alpha_1}{S} + \cdots + \frac{\alpha_d}{S})}{\Gamma(\frac{\alpha_1}{S} + \cdots + \frac{\alpha_d}{S} + n)} \cdot \prod_{i=1}^{d} \frac{\Gamma(\frac{\alpha_i}{S} + k_i)}{\Gamma(\frac{\alpha_i}{S})} \mathbb{1}_{k_1+\cdots+k_d=n}.
\end{aligned}$$

- In order to find $\mathbb{P}(U_i(n) = \alpha_i + Sk_i \ (\forall 1 \geq i \geq d))$, we argue analogously as we did

in Corollary 3.6(ii), and thus:

$$\begin{aligned}\mathbb{P}(U_i(n) = \alpha_i + Sk_i \ (\forall 1 \geq i \geq d)) = \\ &= \binom{n}{k_1, \dots, k_d} \frac{\prod_{i=1}^{d} \alpha_i(\alpha_i + S) \cdots (\alpha_i + (k_i - 1)S)}{\prod_{i=0}^{n-1}(\alpha_1 + \cdots + \alpha_d + iS)} \mathbb{1}_{k_1 + \cdots + k_d = n} \\ &= \binom{n}{k_1, \dots, k_d} \frac{\prod_{i=1}^{d} \frac{\alpha_i}{S}(\frac{\alpha_i}{S} + 1) \cdots (\frac{\alpha_i}{S} + k_i - 1)}{\prod_{i=0}^{n-1}(\frac{\alpha_i + \cdots + \alpha_d}{S} + i)} \mathbb{1}_{k_1 + \cdots + k_d = n} \\ &= \frac{n!}{k_1! \cdots k_d!} \cdot \frac{\prod_{i=1}^{d} \Gamma(\frac{\alpha_i}{S} + k_i)}{\prod_{i=1}^{d} \Gamma(\frac{\alpha_i}{S})} \cdot \frac{\Gamma(\frac{\alpha_1 + \cdots + \alpha_d}{S})}{\Gamma(\frac{\alpha_1 + \cdots + \alpha_d}{S} + n)} \mathbb{1}_{k_1 + \cdots + k_d = n}.\end{aligned}$$

Therefore, $\mathbb{P}(U_i(n) = \alpha_i + Sk_i \ (\forall 1 \geq i \geq d)) = \mathbb{P}(Z_i(n) = \alpha_i + Sk_i \ (\forall 1 \geq i \geq d))$ and this concludes the proof. □

3.2.2 A central limit theorem for the identity case

It is a common inclination to seek a central limit theorem for the fluctuations of a random quantity around its almost sure limit, when dealing with a law of large numbers as in Theorem 3.4. In the identity case, these fluctuations are given by a Gaussian of random variance, function of the almost sure limit V:

Theorem 3.8. *Assume that $(U(n))_{n \geq 0}$ is a Pólya urn of initial composition $(\alpha_1, \dots, \alpha_d)$ and replacement matrix $R = S \cdot \mathrm{Id}$ for some $\alpha_1, \dots, \alpha_d, S \geq 1$. Then, as $n \to \infty$,*

$$\frac{U(n) - nSV}{S\sqrt{n}} \xrightarrow{d} \mathcal{N}(0, \Sigma),$$

where

$$\Sigma = \begin{pmatrix} V_1(1 - V_1) & -V_1V_2 & \dots & -V_1V_d \\ -V_1V_2 & V_2(1 - V_2) & & -V_2V_d \\ \vdots & & \ddots & \vdots \\ -V_1V_d & -V_2V_d & \dots & V_d(1 - V_d) \end{pmatrix}. \tag{3.5}$$

Proof. To prove this, we take advantage of the result found in Theorem 3.7, that is, the equality in law of the processes $U(n)$ and $Z(n)$, defined as $Z(n) = Z(0) + S\sum_{k=1}^{n} X_k$, with $\mathbb{P}(X_k = e_i | V) = V_i$ for every $k \geq 1$, for every $1 \leq i \leq d$. Hence we prove that

$$\frac{Z(n) - nSV}{S\sqrt{n}} \xrightarrow{d} \mathcal{N}(0, \Sigma).$$

By definition of the sequence $X_1, X_2, \dots$, we immediately see that this sequence is, conditionally on V, independent and identically distributed. Also notice that

$$\mathbb{E}[X|V] = e_iV_i \ (\forall 1 \leq i \leq d) = \sum_{i=1}^{d} e_iV_i = V,$$

where X is a copy of X_1. Since the sequence $(X_k)_{k\geq 1}$ is i.i.d., we can apply the multi-dimensional central limit theorem (Theorem 2.9) to get

$$\frac{SX_1 + \cdots + SX_n - nSV}{\sqrt{n}} = \frac{Z(n) - Z(0) - nSV}{\sqrt{n}} \xrightarrow{d} \mathcal{N}(0, \mathrm{Cov}(SX|V)).$$

Let $\Sigma \coloneqq \mathrm{Cov}(X|V)$ since $\mathrm{Cov}(SX|V) = S^2\mathrm{Cov}(X|V) = S^2\Sigma$, we can write the latter convergence in distribution as

$$\frac{Z(n) - Z(0) - nSV}{S\sqrt{n}} \xrightarrow{d} \mathcal{N}(0, \Sigma).$$

Since $\frac{Z(0)}{S\sqrt{n}} \to 0$ almost surely as n goes to infinity, and hence in probability, by Slutsky's theorem (Theorem 2.7), we conclude that

$$\frac{Z(n) - nSV}{S\sqrt{n}} \xrightarrow{d} \mathcal{N}(0, \Sigma).$$

An explicit computation of the covariance matrix, conditionally on V, gives:

$$\begin{aligned}
\Sigma &= \mathbb{E}[(X - \mathbb{E}[X])^T (X - \mathbb{E}[X])|V] \\
&= \sum_{i=1}^{d} \mathbb{P}(X = e_i|V)(e_i - V)^T(e_i - V) \\
&= \sum_{i=1}^{d} V_i(e_i^T e_i - V^T e_i - e_i^T V + V^T V) \\
&= \sum_{i=1}^{d} V_i e_i^T e_i - V^T\left(\sum_{i=1}^{d} V_i e_i\right) - \left(\sum_{i=1}^{d} V_i e_i^T\right)V + V^T V \underbrace{\sum_{i=1}^{d} V_i}_{=1} \\
&= \sum_{i=1}^{d} V_i e_i^T e_i - V^T V - V^T V + V^T V \\
&= \sum_{i=1}^{d} V_i e_i^T e_i - V^T V,
\end{aligned}$$

which corresponds to the matrix illustrated in (3.5) □

3.3 The irreducible matrix case

First, we recall what does it mean for a square matrix to be irreducible and state a well-known result, the Perron-Frobenius Theorem:

Definition 3.9 ([LM95], Definition 4.2.2)**.** A non-negative matrix R is said to be *irreducible* if, for each ordered pair of indices i, j, there exists $n \in \mathbb{N}$ such that $R^n_{i,j} > 0$. As a convention, $R^0 = \mathrm{Id}$, so that the 1×1 matrix $\begin{bmatrix}0\end{bmatrix}$ is irreducible.

Theorem 3.10 (**Perron-Frobenius** ([HJ85], Theorem 8.4.4))**.** *Let $R \neq 0$ be an irreducible matrix. Then R has a positive eigenvector v_R with $\|v_R\|_1 = 1$, and with corresponding eigenvalue $\lambda_R > 0$ (also called spectral radius) that is both geometrically and algebraically simple. If μ is another eigenvalue for R, then $|\mu| < \lambda_R$. Any positive eigenvector for R is a positive multiple of v_R.*

In this section we look at the case, where we consider a Pólya urn process $(U(n))_{n\geq 0}$ with *irreducible* replacement matrix $R = (R_{i,j})_{1\leq i,j\leq d}$. This means that, for every $1 \leq i, j \leq d$, if $U(0) = e_i$, then there is a positive probability to see a ball of colour j at some (finite) time in the urn. Since we assumed the non-diagonal entries of R to be non-negative, there is some $\kappa \geq 0$ such that $R + \kappa\mathrm{Id}$ has non-negative entries. Also, by the binomial theorem, $(R + \kappa \cdot \mathrm{Id})^n = \sum_{i=1}^{d} \binom{n}{i} \kappa^{n-i} R^i$, which is irreducible for every $n \geq 1$. Recall that if μ is a simple eigenvalue of R, then $\mu + \kappa$ is a simple eigenvalue of $R + \kappa \cdot \mathrm{Id}$. Hence, by the Perron-Frobenius theorem (Theorem 3.10), the spectral radius λ_R of $R + \kappa \cdot \mathrm{Id}$, is such that $\lambda_R \geq \mu + \kappa \geq \kappa$ and there is a left-eigenvector v of $R + \kappa \cdot \mathrm{Id}$, whose coefficients are all positive with $\|v\|_1 = 1$. So $\mu = \lambda_R - \kappa > 0$ is the eigenvalue of R with largest real part and $vR = \mu v$. We call μ and v, respectively, the *dominant eigenvalue* and the *dominant left-eigenvector* of R.

3.3.1 A law of large numbers for the irreducible case

The first result presented is again a law of large numbers and is due to Athreya and Karlin [AK68]:

Theorem 3.11. *Fix $d \geq 2$. Let $(U(n))_{n\geq 0}$ be the d-colour Pólya urn of initial composition $(\alpha_1, \ldots, \alpha_d)$ and replacement matrix R. Assume that $\bar{\alpha} := \sum_{i=1}^{d} \alpha_i \geq 1$ and that R is an irreducible matrix. Let $\lambda > 0$ and v be, respectively, the dominant eigenvalue and dominant left-eigenvector of R. Then*

$$\frac{U(n)}{n} \xrightarrow{n\to\infty} \lambda v \quad \text{a.s.}$$

The convergence of $U(n)/n$ towards an almost sure limit as n approaches infinity is the shared feature between Theorems 3.4 (in section 3.2) and 3.11, hence they're both called *law of large numbers*. Nevertheless, these two theorems exhibit significant discrepancies:

- In the irreducible case, the limit of $U(n)/n$ is deterministic (λv), while in the identity case, it's stochastic, represented by a random Dirichlet-distributed vector V.

- Unlike the irreducible scenario, where the limit is independent of the initial composition, in the identity case, the distribution of V depends on the initial distribution.

Remark 3.12. Theorem 3.11 implies that $\|U(n)\|_1/n \to \lambda$ almost surely as $n \to \infty$ because $\|v\|_1 = 1$, by assumption. In other words, the total number of balls in the urn at time n grows as λn as $n \to \infty$.

To prove Theorem 3.11, we embed the process into continuous time, employing a multi-type Galton-Watson process for the embedding, which facilitates precise analysis through martingale theory.
To briefly explain, in general branching processes are stochastic models that describe the evolution of a population over discrete time steps. Each individual in the population gives rise to a random number of offspring in the next generation, and Galton-Watson processes are a specific type of branching process where the number of offspring each individual produces follows a certain probability distribution. For a more detailed exploration of this topic, refer e.g. to [Har02]. Transitioning into continuous time gives more independence, yet necessitates additional effort to translate the results back in discrete time.

3.3.2 Embedding of an urn into continuous time

Given a replacement matrix R and an initial composition $\alpha = (\alpha_1, \ldots, \alpha_d)$, we define the continuous-time, multi-type branching process $(X(t))_{t\geq 0}$ as follows: we let $X(0) = \alpha$, that is, at time $t = 0$, the process starts with α_i particles of type i alive in the system, for $i = 1, \ldots, d$. Each particle reproduces (or "splits") independently from the rest at a constant rate 1 (thus, this corresponds to an exponential distribution of parameter 1). When a particle of type i reproduces, it triggers a reproduction event where $R_{i,j}$ particles of type j are added to the system, for all $1 \leq i, j \leq d$. We call X the *continuous-time urn process* of initial composition α and replacement matrix R.

Proposition 3.13. *Let $\alpha = (\alpha_1, \ldots, \alpha_d)$ and R be a $d \times d$ replacement matrix. Let $(U(n))_{n\geq 0}$ be the Pólya urn of initial composition α and replacement matrix R. Let $(X(t))_{t\geq 0}$ be the continuous-time urn process of initial composition α and replacement matrix R. Let $\tau_0 = 0$ and, for every $n \geq 1$, τ_n be the time when the n-th split in the process $(X(t))_{t\geq 0}$ happens. Then $(U(n))_{n\geq 0} \overset{d}{=} (X(\tau_n))_{n\geq 0}$.*

Proof. The following comes from [AN72], Theorem 2 of chapter V.9, and [CPS11], Lemma 3.1.
For the process $(U(n))_{n\geq 0}$, at each step, if a ball of colour i is drawn, then the drawn ball is put back in the urn, with additional $R_{i,j}$ balls of colour j for every $1 \leq j \leq d$. While for the process $(X(\tau_n))_{n\geq 0}$ the same thing happens: when a particle of type i reproduces, $R_{i,j}$ particles of type j are added to the system, for all $1 \leq j \leq d$. Since both stochastic processes are discrete time Markov chains with stationary transition probabilities, and their initial states and evolution rules are the same ones, the equality in distribution follows. □

3.3.3 A law of large numbers for the continuous-time urn process

In order to prove Theorem 3.11, we first prove a law of large numbers in the continuous-time case, for the continuous-time process X:

Theorem 3.14. *Let X be the continuous-time urn process of initial composition $\alpha = (\alpha_1, \dots, \alpha_d)$ and replacement matrix R. Assume that $\bar{\alpha} = \sum_{i=1}^{d} \alpha_i \geq 1$ and that R is irreducible. Let $\lambda > 0$ and v be, respectively, the dominant eigenvalue and the dominant left-eigenvector of R. Then,*

$$e^{-\lambda t} X(t) \xrightarrow{t\to\infty} Wv \quad \text{a.s.}$$

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

Proof. To establish this, we initially examine a simpler instance: the *Yule process* ([Ros14], section 6.3).
We consider a Yule process $(Y(t))_{t\geq 0}$, its law can be described as follows: for $t = 0$, $Y(0) = 1$ almost surely and if $Y(t) = n$, then the transition to $n + 1$ happens at rate n. To compute the expectation of the Yule process, we denote τ as the first branching event, so we have

$$\begin{aligned}
\mathbb{E}[Y(t)] &= \mathbb{E}[\mathbb{E}[Y(t)|\tau]] \\
&= \mathbb{E}[\mathbb{E}[Y(t)\mathbb{1}_{t<s}|\tau = s]] + \mathbb{E}[\mathbb{E}[Y(t)\mathbb{1}_{t\geq s}|\tau = s]] \\
&= \mathbb{E}[Y(0)\mathbb{1}_{t<\tau}] + \int_0^t \mathbb{E}[Y(t)|\tau = s]\mathbb{P}(\tau \in \mathrm{d}s) \\
&= \mathbb{P}(\tau > t) + \int_0^t \mathbb{E}[Y(t)|\tau = s]\mathbb{P}(\tau \in \mathrm{d}s),
\end{aligned}$$

where the third equality follows since, when $t < s$, for $\tau = s$, the first branching event has not happened yet, so $Y(t) = Y(0) = 1$ almost surely. Notice that $\tau \sim \mathrm{Exp}(1)$, and further the two offsprings that are created evolve independently. For an independent copy Y' of Y, applying the Markov property (Theorem 2.17), it holds:

$$\mathbb{E}[Y(t)|\tau = s] = \mathbb{E}[Y(t-s) + Y'(t-s)] = 2\mathbb{E}[Y(t-s)].$$

Therefore,

$$\mathbb{E}[Y(t)] = e^{-t} + 2\int_0^t e^{-s}\mathbb{E}[Y(t-s)]\,\mathrm{d}s\,.$$

The same approach works for our multi-type branching process. Here, the distinction lies in the outcome of the offspring process: in the Yule process we get two independent copies Y, Y', of the same process, whereas in our case we obtain $1 + \sum_{j=1}^{d} R_{i,j}$ new processes, since in the reproduction event, at time s, a particle of type i splits and

becomes $R_{i,j}+\delta_{i,j}$ particles of type j (where $\delta_{i,j}$ is the Kronecker delta). These particles are denoted by the double-indexed sequence $(X^{(i,j)})_{i,j\geq 1}$, which will be described soon, and which evolve independently starting from time s, so until time t there is $t-s$ time left after applying the Markov property.
Hence, we define by $X^{(i)}$ the urn process of initial composition e_i and replacement matrix R, for all $1 \leq i \leq d$. So for this case $X^{(i)}(0) = e_i$, and we let τ be as in the Yule process case. For all $1 \leq j \leq d$, denote by $(X^{(j,k)})_{k\geq 1}$ the sequence of i.i.d. copies of $X^{(j)}$, and the double-indexed sequence $(X^{(i,j)})_{i,j\geq 1}$ is a sequence of independent processes. Hence,

$$\begin{aligned}
\mathbb{E}[X^{(i)}(t)] &= \mathbb{E}[\mathbb{E}[X^{(i)}(t)\mathbb{1}_{t<s}|\tau = s]] + \mathbb{E}[\mathbb{E}[X^{(i)}(t)\mathbb{1}_{t\geq s}|\tau = s]] \\
&= e_i e^{-t} + \int_0^t e^{-s}\mathbb{E}\left[\sum_{j=1}^{d}\sum_{k=1}^{R_{i,j}+\delta_{i,j}} X^{(j,k)}(t-s)\right] \mathrm{d}s \\
&= e_i e^{-t} + \int_0^t e^{-s}\sum_{j=1}^{d}(R_{i,j}+\delta_{i,j})\mathbb{E}[X^{(j)}(t-s)]\,\mathrm{d}s \\
&= e_i e^{-t} + e^{-t}\int_0^t e^{s}\sum_{j=1}^{d}(R_{i,j}+\delta_{i,j})\mathbb{E}[X^{(j)}(s)]\,\mathrm{d}s\,,
\end{aligned} \tag{3.6}$$

where the last equality follows by the commutativity property of convolutions. So, if we let $u_i(t) \coloneqq \mathbb{E}[X^{(i)}(t)]$ (a horizontal d-dimensional vector) and $u(t)$ be a $d \times d$ matrix whose lines are $u_1(t), \ldots, u_d(t)$, we can write the system of equations, given by equations (3.6) for all $1 \leq i \leq d$, as

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

Differentiating in t yields, $u'(t)e^t + u(t)e^t = e^t u(t)(\mathrm{Id} + R^T)$ which implies $u'(t) = R^T$. So $u(t) = u(0)e^{tR^T} = e^{tR^T}$, since $u(0) = \mathrm{Id}$. Thus, for a general urn process $X(t)$ we have:

$$\mathbb{E}[X(t)] = \sum_{i=1}^{d} X_i(0)u_i(t) = X(0)u(t)^T = X(0)e^{tR}. \tag{3.7}$$

By the definition of continuous-time process X, for all $s, t \geq 0$, we express the process at time $s+t$ as the summation over all components (from 1 up to d), where each component X_i undergoes a reproduction event up to its count $X_i(s)$ at time s, and then evolves according to its own process $X^{(i,j)}(t)$ at time t. This can be summarized by the formula

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

Denote by $\mathcal{F}_t$ the filtration generated by $(X(s))_{s\in[0,t]}$ for all $t \geq 0$. For every $s, t \geq 0$,

we have

$$\begin{aligned}\mathbb{E}[X(s+t)|\mathcal{F}_s] &= \mathbb{E}\left[\sum_{i=1}^{d}\sum_{j=1}^{X_i(s)} X^{(i,j)}(t)\middle|\mathcal{F}_s\right] = \sum_{i=1}^{d}\sum_{j=1}^{X_i(s)}\mathbb{E}[X^{(i,j)}(t)] \\ &= \sum_{i=1}^{d} X_i(s)\mathbb{E}[X^{(i)}(t)] = \sum_{i=1}^{d} X_i(s)e_i e^{tR} \\ &= X(s)e^{tR}\end{aligned}$$

Therefore, for all $s, t \geq 0$,

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

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

In order to prove Theorem 3.14, we follow the proof of Janson ([Jan04], section 9); however to keep things simple, we assume that R is diagonalizable. In the general case, the use of the Jordan decomposition of R becomes necessary, requiring the treatment of projections on each (generalized) eigenspace individually. Dealing with eigenspaces of dimension at least two poses some challenges, which is why we only focus on eigenspaces of dimension 1 here. Under the assumption that R is diagonalizable, other than $v = v_1$, R admits $d-1$ left-eigenvectors $v_2, \ldots, v_d$ associated to the respective $d-1$ eigenvalues $\lambda_2, \ldots, \lambda_d$. Since λ is the dominant eigenvalue of R, we have $\mathrm{Re}(\lambda_i) < \lambda$ for every $2 \leq i \leq d$. Given $x \in \mathbb{R}^d$, we want to express x as a linear combination in the basis $(v_1, \ldots, v_d)$, so that $x = \sum_{i=1}^{d} \pi_i(x)v_i = \sum_{i=1}^{d} \varphi_i(x)$, with $\varphi_i(x) \coloneqq \pi_i(x)v_i$, for every $1 \leq i \leq d$. To do this, we proceed as follows: let $x = \sum_{i=1}^{d} \alpha_i v_i$. Taking the inner product on both sides with v_j, for $j = 1, \ldots, d$, we get

$$\langle x, v_j\rangle = \left\langle \sum_{i=1}^{d} \alpha_i v_i, v_j \right\rangle = \sum_{i=1}^{d} \alpha_i \langle v_i, v_j\rangle = \alpha_j \|v_j\|_2^2,$$

where the last step follows by orthogonality of the basis $(v_i)_{1\leq i\leq d}$. Hence,

$$\pi_j(x) \coloneqq \alpha_j = \frac{\langle x, v_j\rangle}{\|v_j\|_2^2} = \frac{xv_j^T}{v_j v_j^T},$$

for all $1 \leq j \leq d$, and thus we can write $x = \sum_{i=1}^{d} \pi_i(x)v_i = \sum_{i=1}^{d} \varphi_i(x)$.
Before delving into some rates of convergence, we analyse a pair of key properties of the projection π_i.

Lemma 3.16. *For any $1 \leq i \leq d$ and $t \geq 0$, $\pi_i(X(t)e^{-tR}) = e^{-\lambda_i t}\pi_i(X(t))$.*

Proof. By the assumption that R is diagonalizable, we can write $R = K^{-1}DK$, where

$$D = \mathrm{diag}(\lambda_1, \ldots, \lambda_d) \text{ and } K = \begin{bmatrix} v_1 \\ \vdots \\ v_d \end{bmatrix},$$

that is, K is a modal matrix with the eigenvectors of R as rows. Then, it holds that $e^{-tR} = K^{-1}e^{-tD}K$. Notice that, since $(v_i)_{1\leq i\leq d}$ is a basis, we have

$$Kv_i^T = (\langle v_1, v_i\rangle, \dots, \langle v_d, v_i\rangle)^T = e_i^T, \tag{3.8}$$

and so

$$v_i^T = K^{-1}e_i^T. \tag{3.9}$$

Thus,

$$\begin{aligned}\pi_i(X(t)e^{-tR}) &= \frac{\langle X(t)e^{-tR}, v_i\rangle}{\|v_i\|_2^2} = \frac{X(t)e^{-tR}v_i^T}{\|v_i\|_2^2} \\ &= \frac{X(t)K^{-1}e^{-tD}Kv_i^T}{\|v_i\|_2^2} \overset{(3.8)}{=} \frac{X(t)K^{-1}e^{-tD}e_i^T}{\|v_i\|_2^2}.\end{aligned}$$

Now, since $e^{-tD}e_i^T = e^{-\lambda_i t}e_i^T$,

$$\frac{X(t)K^{-1}e^{-tD}e_i^T}{\|v_i\|_2^2} = e^{-\lambda_i t}\frac{X(t)K^{-1}e_i^T}{\|v_i\|_2^2} \overset{(3.9)}{=} e^{-\lambda_i t}\frac{X(t)v_i^T}{\|v_i\|_2^2} = e^{-\lambda_i t}\pi_i(X(t)).$$

Therefore, $\pi_i(X(t)e^{-tR}) = e^{-\lambda_i t}\pi_i(X(t))$, for every $t \geq 0$ and for all $1 \leq i \leq d$. □

Remark 3.17. In many references, where the usual orientation of d-dimensional vectors (with $d \geq 1$) is vertical, the diagonalization of a diagonalizable matrix A is given by $A = PDP^{-1}$, where P is a modal matrix with the eigenvectors of A as columns. As discussed in Notation (2.4), since this thesis considers the transpose orientation, by choosing $K \coloneqq P^T$, we obtain the diagonalization of our replacement matrix R as $R = K^{-1}DK$, where K is a modal matrix with the eigenvectors of R as its rows. Notice that this holds due to the fact that,

$$Av^T = \lambda v^T \iff vA^T = \lambda v,$$

i.e., a right-eigenvector of A is a left-eigenvector of $R \coloneqq A^T$.

Lemma 3.18. *For any $1 \leq i \leq d$ and $t \geq 0$, $e^{-\lambda_i t}\pi_i(X(t))$ is a martingale.*

Proof. By linearity of the inner product and since $X(t)$ is independent of v_i, for any $i = 1, \dots, d$, for $s < t$ we have

$$\mathbb{E}[\pi_i(X(t))|\mathcal{F}_s] = \mathbb{E}\left[\frac{\langle X(t), v_i\rangle}{\|v_i\|_2^2}\middle|\mathcal{F}_s\right] = \frac{\langle \mathbb{E}[X(t)|\mathcal{F}_s], v_i\rangle}{\|v_i\|_2^2} = \pi_i(\mathbb{E}[X(t)|\mathcal{F}_s]),$$

and hence,

$$\begin{aligned}\mathbb{E}[e^{-\lambda_i t}\pi_i(X(t))|\mathcal{F}_s] &= e^{-\lambda_i t}\pi_i(\mathbb{E}[X(t)|\mathcal{F}_s]) = e^{-\lambda_i t}\pi_i(X(s)e^{(t-s)R}) \\ &= e^{-\lambda_i t}\cdot e^{\lambda_i(t-s)}\pi_i(X(s)) = e^{-\lambda_i s}\pi_i(X(s)),\end{aligned}$$

where the second equality follows by Lemma 3.15, and the third equality is due to Lemma 3.16. □

Remark 3.19. Analogously one can show that also $(e^{-\lambda_i t}\varphi_i(X(t)))_{t\geq 0}$ (and thus $(\varphi_i(X(t))e^{-tR})_{t\geq 0}$) is a (vector-valued) martingale.

Now we can prove some rates of convergence for $\pi_i(X(t))$, related to different values of the eigenvalues of R.

Lemma 3.20. *If* $\mathrm{Re}(\lambda_i) > \frac{\lambda}{2}$, *for* $2 \leq i \leq d$, *then*

$$e^{-\lambda_i t}\pi_i(X(t)) \xrightarrow{t\to\infty} W \text{ a.s. and in } L^2.$$

Moreover,

$$e^{-\lambda_i t}\pi_i(X(t)) = \mathbb{E}[W|\mathcal{F}_t], \tag{3.10}$$

almost surely for every $t \geq 0$.

Proof. We aim to show that $e^{-\lambda_i t}\pi_i(X(t))$ is bounded in L^2, so that, by Doob's L^2 martingale convergence theorem (Theorem 2.14), the convergence follows. It is enough to show that $\varphi(X(t))e^{-tR}$ is bounded in L^2, since

$$\begin{aligned}
\mathbb{E}\Big[\big\|\varphi_i(X(t))e^{-tR}\big\|_2^2\Big] &= \mathbb{E}\Big[\big\|\pi_i(X(t))v_i e^{-tR}\big\|_2^2\Big] \\
&= \mathbb{E}\Big[\Big\|e^{-\lambda_i t}\pi_i(X(t))v_i\Big\|_2^2\Big] \\
&= \underbrace{\|v_i\|_2^2}_{\text{const.}} \cdot \mathbb{E}\Big[\Big\|e^{-\lambda_i t}\pi_i(X(t))\Big\|_2^2\Big].
\end{aligned}$$

So, consider

$$Z_k(t) = \sum_{\tau_{k,l}\leq t} \big\|\varphi_l(R_k)e^{-\tau_{k,l}R}\big\|_2^2,$$

where $\tau_{k,l}$ denotes the time of the l-th split of ball of colour k. Notice that computing the quadratic variation of $(\varphi_i(X(t))e^{-tR})_{t\geq 0}$, we have:

$$[(\varphi_i(X(t))e^{-tR})_k]_t = \sum_{\tau_{k,l}\leq t} (\Delta\varphi_i(X(\tau_{k,l}))e^{-\tau_{k,l}R})^2 = \sum_{\tau_{k,l}\leq t} \big\|\Delta\varphi_i(X(\tau_{k,l}))e^{-\tau_{k,l}R}\big\|_2^2$$

Since the jumps occur at times $\tau_{k,l}$, it holds that $\Delta\varphi_i(X(\tau_{k,l})) = \varphi_i(\Delta X(\tau_{k,l}))$. Now, to compute the total quadratic variation of $(\varphi_i(X(t))e^{-tR})_{t\geq 0}$, we must account for contributions from splits of colour k, this gives

$$[\varphi_i(X(t))e^{-tR}]_t = \sum_{k=1}^{d} \sum_{\tau_{k,l}\leq t} \big\|\varphi_i(\Delta X(\tau_{k,l}))e^{-\tau_{k,l}R}\big\|_2^2$$

By definition, each $Z_k(t)$ accounts for the squared contribution from splits of colour k, and thus we have

$$[\varphi_i(X(t))e^{-tR}]_t = \sum_{k=1}^{d} Z_k(t).$$

So, by equation (2.2),

$$\begin{aligned}\mathbb{E}[\left\|\varphi_i(X(t))e^{-tR}\right\|_2^2] &= \mathbb{E}[[\varphi_i(X(t))e^{-tR}]_t] + \mathbb{E}[\|\varphi_i(X(0))\|_2^2] \\ &= \sum_{k=1}^{d} \mathbb{E}[Z_k(t)] + \underbrace{\mathbb{E}[|\varphi_i(X(0))|^2]}_{\text{const.}}.\end{aligned}$$

Therefore, to conclude it is enough to show that $\sup_{t\geq 0}\mathbb{E}[Z_k(t)] < \infty$ for every $1 \leq k \leq d$, so we proceed with the following claim:

Claim. $\left(Z_k(t) - \int_0^t \left\|\varphi_i(R_k)e^{-uR}\right\|_2^2 X_k(u)\,\mathrm{d}u\right)_{t\geq 0}$ *is a martingale of mean zero.*

Proof. By definition of the process $X(t)$, it follows that it is a predictable process, and so we have:

$$\begin{aligned}\mathbb{E}\left[Z_k(t) - \int_0^t \left\|\varphi_i(R_k)e^{-uR}\right\|_2^2 X_k(u)\,\mathrm{d}u \,\middle|\, \mathcal{F}_s\right] &= \mathbb{E}\left[Z_k(s) + \sum_{s<\tau_{k,l}<t} \left\|\varphi_i(R_k)e^{-\tau_{k,l}R}\right\|_2^2 \middle| \mathcal{F}_s\right] \\ &\quad - \mathbb{E}\left[\int_0^s \left\|\varphi_i(R_k)e^{-uR}\right\|_2^2 X_k(u)\,\mathrm{d}u + \int_s^t \left\|\varphi_i(R_k)e^{-uR}\right\|_2^2 X_k(u)\,\mathrm{d}u \,\middle|\, \mathcal{F}_s\right] \\ &= Z_k(s) - \int_0^s \left\|\varphi_i(R_k)e^{-uR}\right\|_2^2 X_k(u)\,\mathrm{d}u \\ &\quad + \mathbb{E}\left[\sum_{s<\tau_{k,l}<t} \left\|\varphi_i(R_k)e^{-\tau_{k,l}R}\right\|_2^2 \middle| \mathcal{F}_s\right] - \int_s^t \left\|\varphi_i(R_k)e^{-uR}\right\|_2^2 X_k(u)\,\mathrm{d}u\,.\end{aligned}$$

Thus, we have a martingale if and only if

$$\mathbb{E}\left[\sum_{s<\tau_{k,l}<t} \left\|\varphi_i(R_k)e^{-\tau_{k,l}R}\right\|_2^2 \middle| \mathcal{F}_s\right] = \int_s^t \left\|\varphi_i(R_k)e^{-uR}\right\|_2^2 X_k(u)\,\mathrm{d}u\,. \tag{3.11}$$

We give an intuitive reasoning on why equality (3.11) holds: by definition, the contribution of $Z_k(t)$ from splits occurring in a small interval $(u, u+\mathrm{d}u)$ is given by $\left\|\varphi_i(R_k)e^{-uR}\right\|_2^2$ per split time, and the expected number of splits in this interval is $X_k(u)\,\mathrm{d}u$. So, for the total expected contribution to $Z_k(t)$ from splits in $(u, u+\mathrm{d}u)$ we have

$$\mathrm{d}Z_k(u) = \left\|\varphi_i(R_k)e^{-uR}\right\|_2^2 X_k(u)\,\mathrm{d}u\,.$$

Integrating and taking conditional expectation on both sides gives us equation (3.11). The martingale is then of mean zero since

$$\mathbb{E}\left[Z_k(t) - \int_0^t \left\|\varphi_i(R_k)e^{-uR}\right\|_2^2 X_k(u)\,\mathrm{d}u\right] = \mathbb{E}[Z_k(0)] = 0.$$

■

By the claim, it follows in particular that,

$$\begin{aligned}\mathbb{E}[Z_k(t)] &= \int_0^t \left\|\varphi_i(R_k)e^{-sR}\right\|_2^2 \mathbb{E}[X_k(s)]\,\mathrm{d}s \\ &= \int_0^t \left\|e^{-s\lambda_i}\varphi_i(R_k)\right\|_2^2 \mathbb{E}[X_k(s)]\,\mathrm{d}s \\ &= \int_0^t e^{-2s\lambda_i}\|\varphi_i(R_k)\|_2^2 \mathbb{E}[X_k(s)]\,\mathrm{d}s\,.\end{aligned}$$

Now, recall that, by equation (3.7), we have $\mathbb{E}[X(t)] = X(0)e^{tR}$, and so $\mathbb{E}[X_k(s)] = (X(0)e^{sR})_k = O(e^{\lambda s})$. Thus, the integrand satisfies

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

which is integrable, since by assumption $2\,\mathrm{Re}(\lambda_i) > \lambda$. This implies that

$$\sup_{t\geq 0} \mathbb{E}[Z_k(t)] < \infty,$$

which concludes the proof.
The relation (3.10) follows directly from the fact that $e^{-\lambda_i t}\pi_i(X(t))$ is a martingale (Lemma 3.18) and from the martingale convergence theorem (Theorem 2.12). □

Lemma 3.21. *If* $\mathrm{Re}(\lambda_i) > \frac{\lambda}{2}$ *for every* $1 \leq i \leq d$, *then* $\mathbb{P}(Wv_i = 0) < 1$.

Proof. By contradiction, suppose that $Wv_i = 0$ almost surely. By Lemma 3.20 we have $\mathbb{E}[W|\mathcal{F}_t] = \varphi_i(X(t)e^{-tR})$ for every $t \geq 0$, which then implies that $\pi_i(X(t)e^{-tR}) = 0$, $\forall t \geq 0$. But since, by Lemma 3.16, $\pi_i(X(t)e^{-tR}) = e^{-t\lambda_i}\pi_i(X(t))$ for every $t \geq 0$, the latter implies that $\pi_i(X(t)) = 0$. This holds almost surely, simultaneously for all $t \in \mathbb{Q}$, and using right-continuity, we get that $(\pi_i(X(t)))_{t\geq 0}$ is almost surely the constant, null function. Then, it also holds that $\varphi_i(\Delta X(t)) = 0$, $\forall t \geq 0$, and therefore $\varphi_i(R_k) = 0$, $\forall 1 \leq k \leq d$. This means that, $\varphi_i R = 0$, which implies $\lambda_i \varphi_i = 0$, and thus $\lambda_i = 0$, which is impossible since, by assumption, $\mathrm{Re}(\lambda_i) > \frac{\lambda}{2} > 0$. □

We look now at the case, where $\mathrm{Re}(\lambda_i) \leq \frac{\lambda}{2}$:

Lemma 3.22. *If* $\mathrm{Re}(\lambda_i) \leq \frac{\lambda}{2}$, *for* $2 \leq i \leq d$, *then* $\pi_i(X(t)) = o(e^{\lambda t})$ *almost surely as* $t \to \infty$.

Proof. Fix λ_i, for $2 \leq i \leq d$, let $\varepsilon > 0$ and $\delta = \lambda - \mathrm{Re}(\lambda_i) > 0$. Further, let

$$A_n = \left\{ \sup_{t\in[n-1,n]} \left|e^{-\lambda t}\pi_i(X(t))\right| > \varepsilon \right\}.$$

Then, on A_n, for some $t \in [n-1, n]$, we have

$$\varepsilon e^{\lambda t} < |\pi_i(X(t))| = \left|e^{\lambda_i t}\pi_i(X(t)e^{-tR})\right| = e^{\mathrm{Re}(\lambda_i)t}\left|\pi_i(X(t)e^{-tR})\right|,$$

by Lemma 3.16. Thus, $\left|\pi_i(X(t)e^{-tR})\right| > \varepsilon e^{\delta t}$. Using Doob's maximal inequality (Theorem 2.15), we get

$$\begin{aligned}\mathbb{P}(A_n) &\leq \mathbb{P}\bigg(\sup_{t\leq n}\left|\pi_i(X(t)e^{-tR})\right| \geq \varepsilon e^{(n-1)\delta}\bigg)\\ &\leq \frac{1}{\varepsilon^2 e^{2n\delta}}\mathbb{E}\Big[\left|\pi_i(X(n)e^{-nR}\right|^2\Big]\\ &\leq \frac{1}{\varepsilon^2 e^{2n\delta}}e^{\max(0,2\delta-\lambda)n},\end{aligned}$$

where the last inequality follows by the fact that, in the proof of Lemma 3.20 we showed that, $\mathbb{E}\big[\left|\pi_i(X(t)e^{-tR})\right|^2\big] = O\big(e^{t(\lambda-2\operatorname{Re}(\lambda_i))}\big) = O\big(e^{t(2\delta-\lambda)}\big)$.
Summing over n, the last term is finite, hence, by the Borel-Cantelli Lemma (Theorem 2.5), we get $\limsup_{t\to\infty}\left|e^{-\lambda t}\pi_i(X(t))\right| \leq \varepsilon$. Since ε is arbitrary, this means that,

$$e^{-\lambda t}\pi_i(X(t)) \xrightarrow{t\to\infty} 0 \quad \text{a.s.}$$

for $2 \leq i \leq d$, in other words, $\pi_i(X(t)) = o(e^{\lambda t})$. □

Now we have what we need to prove Theorem 3.14, and consequently Theorem 3.11.

Proof of Theorem 3.14. Using the expression for π_i, we can write

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

Then by Lemma 3.20 and Lemma 3.22, $e^{-\lambda t}\pi_i(X(t)) = o(1)$ for every $2 \leq i \leq d$, and therefore

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

as desired. □

We can now use the connection between the Pólya urn $(U(n))_{n\geq 0}$ and the continuous-time urn process $(X(t))_{t\geq 0}$ made in Proposition 3.13 to prove Theorem 3.11, which then follows by Theorem 3.14.

Proof of Theorem 3.11. Notice that, since $\|v\|_1 = 1$, by Theorem 3.14, we have

$$e^{-\lambda t}\|X(t)\|_1 \xrightarrow{t\to\infty} W \tag{3.12}$$

almost surely. This implies that

$$\frac{X(t)}{\|X(t)\|_1} \xrightarrow{t\to\infty} v \text{ a.s.} \tag{3.13}$$

and therefore, since $\tau_n \to \infty$ as $n \to \infty$, by Proposition 3.13,

$$\frac{U(n)}{\|U(n)\|_1} \xrightarrow{n\to\infty} v \text{ a.s.} \tag{3.14}$$

Following [AN72], section 7 of chapter V, for all $1 \leq i \leq d$ and $k \geq 1$, we define $\delta_k^{(i)}$ as the indicator that the k-th split in X is from a particle of colour i. For all $k \geq 1$, let $\mathcal{G}_n$ denote the sub-σ-algebra generated by the process X until time τ_k. Then, by definition, we have

$$\mathbb{P}(\delta_k^{(i)} = 1|\mathcal{G}_n) = \frac{X_i(\tau_k^-)}{\left\|X(\tau_k^-)\right\|_1},$$

which by (3.13) converges almost surely to v_i as k goes to infinity. Now we define $N_i(n)$ as the number of splits triggered by a particle of colour i within the first n splits, and hence by definition $N_i(n) = \sum_{i=1}^n \delta_k^{(i)}$. So, for every $n \geq 0$, we have

$$\frac{N_i(n)}{n} = \frac{1}{n}\sum_{k=1}^n \delta_k^{(i)} = \frac{1}{n}\sum_{k=1}^n \mathbb{P}(\delta_k^{(i)} = 1|\mathcal{G}_k) + \frac{1}{n}\sum_{k=1}^n \left(\delta_k^{(i)} - \mathbb{P}(\delta_k^{(i)} = 1|\mathcal{G}_k)\right). \tag{3.15}$$

By a strong law of large numbers for martingales due to Lévy ([Lév37], pp. 250 as mentioned in [AN72], Lemma 1 of V.7), the second summand goes to zero as $n \to \infty$. Now we recall a result from analysis, the Stolz-Cesàro Theorem (Theorem 2.7.2 of [CN14]):

Theorem 3.23 (Stolz-Cesàro Theorem). *Let $(a_n)_{n\geq 0}$ and $(b_n)_{n\geq 0}$ be two sequences of real numbers such that $(b_n)_{n\geq 0}$ is strictly increasing to ∞ and*

$$\lim_{n\to\infty} \frac{a_{n+1} - a_n}{b_{n+1} - b_n} = l \text{ (possibly in } \mathbb{R} \cup \{+\infty\}).$$

Then the limit $\lim_{n\to\infty} \frac{a_n}{b_n}$ also exists and equals l.

Thus, by Theorem 3.23, choosing $a_n = \mathbb{P}(\delta_k^{(i)} = 1|\mathcal{G}_n)$ and $b_n = n$, the first term of (3.15) converges almost surely to v_i. Therefore, for all $1 \leq i \leq d$,

$$\frac{N_i(n)}{n} \xrightarrow{n\to\infty} v_i \text{ a.s..}$$

Then, for every $n \geq 0$,

$$\frac{\|X(\tau_n)\|_1}{n} = \frac{1}{n}\sum_{i=1}^d \sum_{j=1}^n \delta_j^{(i)} \|R_i\|_1 = \sum_{i=1}^d \|R_i\|_1 \frac{N_i(n)}{n} \xrightarrow{n\to\infty} \sum_{i=1}^d v_i \|R_i\|_1 \tag{3.16}$$

almost surely. Now we notice that

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

Hence, combining equations (3.13), (3.14), (3.16) and Proposition 3.13, we get that

$$\frac{U(n)}{n} \xrightarrow{n\to\infty} \lambda v \quad \text{a.s.}$$

□

3.3.4 A central limit theorem for the irreducible case

After establishing the result of almost sure convergence, it is natural to examine the fluctuations of $U(n)/n$ around its limit λv. The results in this section are based on the work of Athreya et al. ([AN72], [Ath69a]) and Janson ([Jan04]). Our aim is to prove the following theorem, under the assumption that R is diagonalizable.

Theorem 3.24. *Under the assumption of Theorem 3.11, let* $\sigma = \max_\mu\{\mathrm{Re}(\mu)/\lambda\}$, *where the maximum is taken over all the eigenvalues of* R *and* $\mu \neq \lambda$.

(i) *If* $\sigma < \frac{1}{2}$, *then, as* $n \to \infty$,

$$\frac{U(n) - n\lambda v}{\sqrt{n}} \xrightarrow{d} \mathcal{N}(0, \Sigma),$$

where Σ *is known and depends on* R.

(ii) *If* $\sigma = \frac{1}{2}$, *then, as* $n \to \infty$,

$$\frac{U(n) - n\lambda v}{\sqrt{n(\log(n))}} \xrightarrow{d} \mathcal{N}(0, \Sigma),$$

where Σ *is known and depends on* R.

(iii) *If* $\sigma > \frac{1}{2}$, *then, almost surely as* $n \to \infty$,

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

To achieve this, we will first focus on the case of *small eigenvalues* (when $\mathrm{Re}(\lambda_i) < \lambda/2$), followed by the case $\mathrm{Re}(\lambda_i) = \lambda/2$, and finally, the case of *big eigenvalues* (when $\mathrm{Re}(\lambda_i) > \lambda/2$). Before proceeding, it is useful to observe the following results regarding the n-th splitting times τ_n. Without loss of generality, we assume $\max_{\lambda_i \neq \lambda}(\mathrm{Re}(\lambda_i)) = \mathrm{Re}(\lambda_2)$.

Lemma 3.25. *Let* τ_n *be the* n*-th splitting time in the process* $(X(t))_{t\geq 0}$, *and let* W *be the random variable resulting from the convergence of Theorem 3.14. Then, as* $n \to \infty$,

$$ne^{-\lambda\tau_n} \to \frac{W}{\lambda} \quad a.s.$$

Proof. This is just an observation of what we have seen in the proof of Theorem 3.11. Indeed, by relation (3.12), we have, almost surely,

$$e^{-\lambda\tau_n}\|X(\tau_n)\|_1 \xrightarrow{n\to\infty} W,$$

and by (3.16), we know that $\|X(\tau_n)\|_1 \approx \lambda n$ a.s. Hence, we get $ne^{-\lambda\tau_n}\lambda \xrightarrow{n\to\infty} W$ a.s. □

Lemma 3.26. *Under the assumptions of Lemma 3.25,*

$$\tau_n - \frac{1}{\lambda}\log(n) \xrightarrow{n\to\infty} -\frac{1}{\lambda}(\log(W) - \log(\lambda)) \quad \text{a.s.}$$

Proof. This follows directly by Lemma 3.25, taking logarithms. □

Lemma 3.27. *As $t \to \infty$, $e^{-\frac{\lambda t}{2}} \sup_{s\le t} \|\Delta X(s)\|_2 \to 0$ almost surely.*

Proof. Let $M > 0$ and define

$$Z(t) \coloneqq \sum_{i=1}^{d} Z_k(t) = \sum_{i=1}^{d} \sum_{\tau_{k,l}\le t} \mathbb{1}_{\Delta X(\tau_{k,l})>M}$$

as the number of jumps larger than M until time t, where, as before, $\tau_{k,l}$ denotes the time of the l-th split of ball of colour k. We have,

$$\mathbb{P}\Big(\sup_{s\le t} \|\Delta X(s)\|_2 > M\Big) = \mathbb{P}\big(Z(t) \ge 1\big) \le \mathbb{E}\big[Z(t)\big].$$

Arguing analogously as in the proof of Lemma 3.20, we get that

$$\mathbb{E}\big[Z_k(t)\big] = \int_0^t \mathbb{P}\Big(\|\varphi_i(R_k)\|_2^2 > M\Big)\mathbb{E}\big[X_k(s)\big]\,\mathrm{d}s = \mathbb{P}\Big(\|\varphi_i(R_k)\|_2^2 > M\Big)\cdot O\big(e^{\lambda t}\big).$$

Now, let $\varepsilon > 0$, and let

$$B_n = \Big\{ \sup_{t\in[n,n+1]} e^{-\frac{\lambda t}{2}} \sup_{s\le t} \|\Delta X(s)\| > \varepsilon \Big\}.$$

Taking $M = \varepsilon e^{\frac{\lambda n}{2}}$, yields

$$\begin{aligned}
\mathbb{P}(B_n) &\le \mathbb{P}\Big(\sup_{s\le n+1} \Delta X(s) > M\Big) \le \mathbb{E}[Z(n+1)] \\
&\le C\sum_{i=1}^{d} \mathbb{P}\big(\|\varphi_i(R_k)\|_2^2 > M\big)e^{\lambda(n+1)} \\
&= C\sum_{i=1}^{d} \mathbb{P}\left(\frac{\|\varphi_i(R_k)\|_2^2}{\varepsilon^2} > e^{\lambda n}\right)e^{\lambda(n+1)},
\end{aligned}$$

for some constant C. Then,

$$\begin{aligned}
\sum_{n\ge1} \mathbb{P}(B_n) &\le C_1 \sum_{i=1}^{d} \mathbb{E}\left[\sum_{n\ge1} e^{\lambda(n+1)} \mathbb{1}_{e^{\lambda n}<\varepsilon^{-2}\|\varphi_i(R_k)\|_2^4}\right] \\
&\le C_1 \sum_{i=1}^{d} \mathbb{E}\left[\sum_{n\le\lambda^{-1}\log\left(\varepsilon^{-2}\|\varphi_i(R_k)\|_2^4\right)} e^{\lambda}\frac{\|\varphi_i(R_k)\|_2^4}{\varepsilon^2}\right] \\
&\le C_2 \sum_{i=1}^{d} \mathbb{E}\left[\frac{\|\varphi_i(R_k)\|_2^4}{\varepsilon^2}\right] < \infty,
\end{aligned}$$

for some constants C_1, C_2. Notice that, in the last inequality we used the fact that, for a real x, it holds $\mathbb{1}_{x^2\ge1} \le x^2$.
The result then follows by the Borel-Cantelli Lemma (Theorem 2.5). □

Lemma 3.28. *As $n \to \infty$, $\sup_{s \leq \tau_n} \frac{\|\Delta X(s)\|_2}{\sqrt{n}} \to 0$ almost surely.*

Proof. Since $\tau_n \to \infty$ as $n \to \infty$, by Lemma 3.27,

$$e^{-\frac{\lambda \tau_n}{2}} \sup_{s \leq \tau_n} \|\Delta X(s)\|_2 \to 0 \quad \text{a.s.}$$

and by Lemma 3.25, we have

$$\frac{e^{\frac{\lambda \tau_n}{2}}}{\sqrt{n}} \to \frac{\sqrt{\lambda}}{\sqrt{W}} \quad \text{a.s.}$$

By multiplication, we get the result. □

We now define the stopping time, where the process is stopped as soon as either we reach a given total number of balls, or a given number of balls of a certain colour. This will help us in proving Theorem 3.24. So, formally, this time is defined as

$$\tau_b(z) = \min \big\{ t \geq 0 : \langle X(t), b \rangle \geq z \big\}, \tag{3.17}$$

for a fixed $b \in \mathbb{R}^d$ and $z \geq 0$. We also assume that $\langle b, v \rangle > 0$.

Remark 3.29. Lemmas 3.25-3.28 hold also taking $\tau_b(z)$ instead of τ_n, in this case, the resulting limit, as $z \to \infty$, has $\lambda = \langle b, v \rangle^{-1}$. We will later see, that this choice is not casual, e.g. in proving Theorem 3.24.

Now, we define a projection similar to $\varphi_i(x)$: let $P_i = u_i^T v_i$ be a $d \times d$ matrix, where u_i is the right-eigenvector of R with associated eigenvalue λ_i, and for $x \in \mathbb{R}^d$, let $P_i(x) = x u_i^T v_i$. Since R is diagonalizable, it follows that, the bases of eigenvectors (respectively left- and right-) $(v_i)_{1 \leq i \leq d}$ and $(u_i)_{1 \leq i \leq d}$ are bi-orthogonal, meaning that $u_i v_j^T = \delta_{i,j}$, where $\delta_{i,j}$ is the Kronecker's delta. To see this, it is enough to notice that, since $R = K^{-1} D K$ (as we did in the proof of Lemma 3.16), equivalently

$$KR = DK \quad \text{and} \quad RK^{-1} = K^{-1} D,$$

that is, the rows of K are the left-eigenvectors of R, and respectively, the columns of K^{-1} are the right-eigenvectors of R. Of course $K^{-1}K = \mathrm{Id}$, hence the bi-orthogonality follows. Moreover, similarly as we did in section 3.3.3, let $x = \sum_{i=1}^d \alpha_i v_i$, and notice that,

$$\sum_{j=1}^d P_j(x) = \sum_{j=1}^d \sum_{i=1}^d \alpha_i v_i u_j^T v_j = \sum_{j=1}^d \alpha_j v_j = x.$$

Therefore, we found a projection $P_i(x)$ similar to $\varphi_i(x)$, in the sense that we can express an element x as a sum of these projections:

$$x = \sum_{i=1}^d \varphi_i(x) = \sum_{i=1}^d P_i(x).$$

In particular, we also have that $\sum_{i=1}^d P_i = \sum_{i=1}^d u_i^T v_i = \mathrm{Id}$.
With these new notions, we are ready to state the following result.

Proposition 3.30. *As $z \to \infty$, almost surely,*

$$\textit{(i)}\ \frac{1}{\sqrt{z}}\left(P_1(X(\tau_b(z))) - \frac{z}{\langle b, v\rangle}v + \sum_{i=2}^{d}\frac{\langle b, P_i(X(\tau_b(z)))\rangle}{\langle b, v\rangle}v\right) \to 0,$$

$$\textit{(ii)}\ \frac{1}{\sqrt{z}}\left(X(\tau_b(z)) - \frac{z}{\langle b, v\rangle}v - \sum_{i=2}^{d}P_i(X(\tau_b(z)))\left(\mathrm{Id} - \frac{b^T v}{\langle b, v\rangle}\right)\right) \to 0$$

Proof. For the first result, notice that, since $\tau_b(z)$ is the first time $\langle b, X(t)\rangle$ is larger or equal to z, we have $\langle b, X(\tau_b^-(z))\rangle \leq z$. So, $\langle b, X(\tau_b(z))\rangle - z \leq \langle b, \Delta X(\tau_b(z))\rangle$. By the Cauchy-Schwarz inequality (Theorem 2.1, choosing $p = q = 2$), we get

$$|\langle b, X(\tau_b(z))\rangle - z| \leq \|b\|_2 \|\Delta X(\tau_b(z))\|_2.$$

So, by Lemma 3.28, it follows that $z^{-1/2}(\langle b, X(\tau_b(z))\rangle - z) \to 0$ as $z \to \infty$ almost surely. Multiplying this by $v/\langle b, v\rangle$, writing $X(\tau_b(z)) = \sum_{i=1}^{d} P_i(X(\tau_b(z)))$, and noticing that, for every $v_i \in \mathrm{im}(P_1)$, i.e. $v_i \in \mathrm{span}(v)$, we can write $v_i = \alpha v$, for some scalar α. Then,

$$\frac{\langle b, v_i\rangle}{\langle b, v\rangle}v = \frac{\langle b, \alpha v\rangle}{\langle b, v\rangle}v = \alpha v = v_i.$$

Part (ii) follows by (i) and the fact that $X(\tau_b(z)) = \sum_{i=1}^{d} P_i(X(\tau_b(z)))$. □

We can now derive our Theorem 3.24 case by case.

3.3.4.1 Small eigenvalues

Throughout this section we assume that $\mathrm{Re}(\lambda_i) < \lambda/2$ for $i = 2, \ldots, d$. We start with some results proved by Athreya et al. and then make a connection with the paper of Janson. Consider the process $Y(t) = l_1 \,\mathrm{Re}(\langle v_i, X(t)\rangle) + l_2 \,\mathrm{Im}(\langle v_i, X(t)\rangle)$, for $l_1, l_2 \in \mathbb{R}$. Then we have the following result:

Theorem 3.31. *Let $\sigma^2 = \lim_{t\to\infty}\sum_{i=1}^{d} v_{1,i}e^{-\lambda t}\mathbb{E}\big[|Y(t)|^2\big]$, and denote by $\Phi(x)$ the cumulative distribution function of the standard normal distribution, i.e.*

$$\Phi(x) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x} e^{-\frac{t^2}{2}}\,\mathrm{d}t\,.$$

Then,

$$\mathbb{P}\left(0 < x_1 \leq W \leq x_2 < \infty, \frac{Y(t)}{\sqrt{\langle v, X(t)\rangle}} \leq x\right) \xrightarrow{t\to\infty} \mathbb{P}\big(0 < x_1 \leq W \leq x_2 < \infty\big)\Phi\Big(\frac{x}{\sigma}\Big).$$

Remark 3.32. The so called Cramer-Wold device (cf. Theorem 10.5 of [Gut13]), which says that a sequence of random vectors converges in distribution if, and only if, the distribution of every linear combination of its components also converges, implies that, by the arbitrariness of l_1 and l_2, the joint distribution of

$$\left(\frac{\mathrm{Re}(\langle v_i, X(t)\rangle)}{\sqrt{\langle v, X(t)\rangle}}, \frac{\mathrm{Im}(\langle v_i, X(t)\rangle)}{\sqrt{\langle v, X(t)\rangle}}\right)$$

approaches the bivariate normal. This is enough, to show that the original process $X(t)$ divided by $\sqrt{\langle v, X(t)\rangle}$ converges in law to a normal distribution, and this is why we consider the process $Y(t)$.

As we noticed in the proof of Lemma 3.15, we can write

$$Y(t+s) = \sum_{i=1}^{d} \sum_{j=1}^{X_i(t)} Z_{i,j}(s) + \tilde{Y}(t,s),$$

where

$$\begin{aligned} Z_{i,j}(s) &= l_1 \operatorname{Re}(\langle v, X^{(i,j)}(s)\rangle - e^{\lambda s} v_1) + l_2 \operatorname{Im}(\langle v, X^{(i,j)}(s)\rangle - e^{\lambda s} v_1), \\ \tilde{Y}(t,s) &= l_1 \operatorname{Re}(e^{\lambda s}\langle v, X(t)\rangle) + l_2 \operatorname{Im}(e^{\lambda s}\langle v, X(t)\rangle). \end{aligned}$$

Dividing both sides by $\sqrt{\langle v, X(t+s)\rangle}$, we get

$$\begin{aligned} \frac{Y(t+s)}{\sqrt{\langle v, X(t+s)\rangle}} = &\underbrace{\sum_{i=1}^{d} \frac{\sqrt{v_i}}{\sqrt{X_i(t)}} \sum_{j=1}^{X_i(t)} Z_{i,j}(s) e^{-\frac{\lambda s}{2}}}_{A_1(t,s)=:A_1} \\ &+ \underbrace{\sum_{i=1}^{d} \left(\sqrt{\frac{X_i(t)e^{-\lambda t}}{\langle v, X(t+s)\rangle e^{-\lambda(t+s)}}} - \sqrt{v_i} \right) \frac{1}{\sqrt{X_i(t)}} \left(\sum_{j=1}^{X_i(t)} Z_{i,j}(s) e^{-\frac{\lambda s}{2}} \right)}_{A_2(t,s)=:A_2} \\ &+ \underbrace{\frac{\tilde{Y}(t,s)}{\sqrt{\langle v, X(t+s)\rangle}}}_{A_3(t,s)=:A_3} \end{aligned}$$

We proceed in proving Theorem 3.31, by showing some properties of A_1, A_2 and A_3 respectively. Let $I = [x_1, x_2]$ for $x_1 > 0, x_2 < \infty$.

Lemma 3.33. *Let $\eta > 0, \delta > 0$. Then there exists $s_0 := s(\eta, \delta)$ such that for every $s \geq s_0$,*

$$\sup_t \mathbb{P}(W \in I, |A_3| > \eta) \leq \delta.$$

Proof. First, we have

$$\begin{aligned} \mathbb{P}\big(W \in I, |A_3| > \eta\big) \leq &\mathbb{P}\bigg(W \in I, |A_3| > \eta, \Big|\langle v, X(t+s)\rangle e^{-\lambda(t+s)} - W\Big| \leq \varepsilon\bigg) \\ &+ \mathbb{P}\bigg(W \in I, \Big|\langle v, X(t+s)\rangle e^{-\lambda(t+s)} - W\Big| > \varepsilon\bigg), \end{aligned}$$

where $\varepsilon = \frac{x_1}{2}$. But, by Lemma 3.20, $\langle v, X(t+s)\rangle e^{-\lambda(t+s)} \to W$ almost surely, hence, given $\delta > 0$, there is $s_0' = s(\delta)$ such that for every $s \geq s_0'$,

$$\sup_t \mathbb{P}\bigg(\Big|\langle v, X(t+s)\rangle e^{-\lambda(t+s)} - W\Big| > \varepsilon\bigg) < \frac{\delta}{2}.$$

Now, notice that,

$$\begin{aligned}\mathbb{P}\bigg(W \in I, |A_3| > \eta, \Big|\langle v, X(t+s)\rangle e^{-\lambda(t+s)} - W\Big| \le \varepsilon\bigg) &\le \mathbb{P}\bigg(\Big|\tilde{Y}(t,s)e^{-\lambda\frac{(t+s)}{2}}\Big| > \eta\sqrt{\frac{x_1}{2}}\bigg)\\ &\le \frac{2e^{-\lambda(t+s)}}{\eta^2 x_1}\mathbb{E}\Big[\Big|\tilde{Y}(t+s)\Big|^2\Big]\\ &\le \frac{2(l_1^2+l_2^2)}{\eta^2 x_1}\mathbb{E}\Big[|\langle v, X(t)\rangle|^2\Big]e^{-\lambda t}e^{(2\operatorname{Re}(\lambda_i)-\lambda)s},\end{aligned}$$

where the last term converges, as we have shown in the proof of Lemma 3.20. Also notice, that the second inequality follows by the Markov inequality (Theorem 2.3). So, we can conclude that,

$$\mathbb{P}\bigg(W \in I, |A_3| > \eta, \Big|\langle v, X(t+s)\rangle e^{-\lambda(t+s)} - W\Big| \le \varepsilon\bigg) \le \frac{2(l_1^2+l_2^2)}{\eta^2 x_1}Ce^{(2\operatorname{Re}(\lambda_i)-\lambda)s},$$

and we can choose $s_0'' \coloneqq s(\delta,\eta)$ such that for every $s > s_0''$,

$$\frac{2(l_1^2+l_2^2)}{\eta^2 x_1}Ce^{(2\operatorname{Re}(\lambda_i)-\lambda)s_0''} < \frac{\delta}{2}.$$

The proof is complete by taking $s_0 = \max(s_0', s_0'')$. □

Lemma 3.34. *For a fixed s and any $\eta > 0$, $\mathbb{P}(W \in I, |A_2| > \eta) \xrightarrow{t\to\infty} 0$.*

Proof. We can write $A_2(t,s) = \sum_{i=1}^d C_i(t,s)D_i(t,s)$, where

$$\begin{aligned}C_i(t,s) &= \sqrt{\frac{X_i(t)e^{-\lambda t}}{\langle v, X(t+s)\rangle e^{-\lambda(t+s)}}} - \sqrt{v_i},\\ D_i(t,s) &= \frac{1}{\sqrt{X_i(t)}}\sum_{j=1}^{X_i(t)} Z_{i,j}(s)e^{-\frac{\lambda s}{2}}.\end{aligned}$$

It is enough to show, that for each i-th component, $\mathbb{P}(W \in I, |C_iD_i| > \eta) \xrightarrow{t\to\infty} 0$. As before,

$$\begin{aligned}\mathbb{P}(W \in I, |C_iD_i| > \eta) \le{}& \mathbb{P}(W \in I, |C_iD_i| > \eta, |D_i| \le M)\\ &+ \mathbb{P}(W \in I, |C_iD_i| > \eta, |D_i| > M).\end{aligned}$$

As discussed in the proof of Lemma 3.15, conditionally on $X_i(t)$, the $Z_{i,j}(s)e^{-\frac{\lambda s}{2}}$ are independent and identically distributed, moreover of mean 0 and finite variance. Hence,

$$\mathbb{E}\big[|D_i|^2\big] = \mathbb{E}\bigg[\Big|Z_{i,j}(s)e^{-\frac{\lambda s}{2}}\Big|^2\bigg] \le C,$$

for some positive C. Now, let $\varepsilon > 0$ and choose M large enough such that $\frac{C}{M^2} < \varepsilon$. Then,

$$\mathbb{P}(W \in I, |C_iD_i| > \eta, |D_i| \le M) \le \mathbb{P}\bigg(W \in I, |C_i| > \frac{\eta}{M}\bigg).$$

Since $x_1 > 0$, combined with Theorem 3.14 and Lemma 3.20, we get that

$$\mathbb{P}\left(W \in I, |C_i| > \frac{\eta}{M}\right) \xrightarrow{t\to\infty} 0.$$

Therefore, for an arbitrary ε,

$$\sup_t \mathbb{P}(W \in I, |C_i D_i| > \eta) \leq \varepsilon.$$

□

Lemma 3.35. *For a fixed s_0 and x, $\mathbb{P}(W \in I, A_1(t, s_0) \leq x) \xrightarrow{t\to\infty} \mathbb{P}(W \in I)\Phi\big(\frac{x}{\sigma(s_0)}\big)$.*

Proof. This follows directly by the following two claims:

Claim 1. *Let $d(x, I) = \inf_{y\in I} |x - y|$ and set $\eta_t = \inf_{t_\eta \leq t}$. Let η, δ be arbitrary, $s = s_0(\eta, \delta)$, where s_0 is as in Lemma 3.33. Let $\varepsilon > 0$, then there exists t_0 such that for all $t \geq t_0$, and for any y,*

$$\Big|\mathbb{P}(W \in I, A_1(t, s_0) \leq y) - \mathbb{P}(d(\langle v, X(t)\rangle e^{-\lambda t}, I) \leq \eta_t, A_1(t, s_0) \leq y)\Big| < 3\varepsilon. \tag{3.18}$$

Claim 2. *Consider s_0 and y as in Claim 1, then*

$$\mathbb{P}(d(\langle v, X(t)\rangle e^{-\lambda t}, I) \leq \eta_t, A_1(t, s_0) \leq y) \xrightarrow{t\to\infty} \mathbb{P}(W \in I)\Phi\left(\frac{y}{\sigma(s_0)}\right)$$

Proof of Claim 1. Let $A\Delta B$ denote the symmetric difference between A and B. Let $\varepsilon > 0$. Then, for every $\eta > 0$, there exists t_η such that, for all $t \geq t_\eta$,

$$\mathbb{P}\left(\Big|d(\langle v, X(t)\rangle e^{-\lambda t}, I) - d(W, I)\Big| > \eta\right) \leq \varepsilon.$$

For all $t > 0$, we set $\eta_t = \inf\{\eta : t_\eta \leq t\}$, then let A denote the event $W \in I$ and B the event $d(\langle v, X(t)\rangle e^{-\lambda t}, I) \leq \eta_t$. So, $\mathbb{P}(A\Delta B) = \mathbb{P}(A \cap B^{\mathsf{c}}) + \mathbb{P}(A^{\mathsf{c}} \cap B)$. By definition of B, $\mathbb{P}(A \cap B^{\mathsf{c}}) \leq \mathbb{P}(\big|d(\langle v, X(t)\rangle e^{-\lambda t}, I) - d(W, I)\big| > \eta_t) \leq \varepsilon$, but

$$\mathbb{P}(A^{\mathsf{c}} \cap B) = \mathbb{P}(A^{\mathsf{c}} \cap B, d(W, I) \leq r) + \mathbb{P}(A^{\mathsf{c}} \cap B, d(W, I) > r) \tag{3.19}$$

for an arbitrary r. Next, $\mathbb{P}(A^{\mathsf{c}} \cap B, d(W, I) \leq r) \leq \mathbb{P}(0 < d(W, I) \leq r)$. By assumption, W is continuous on $(0, \infty)$, so there is some $r > 0$ such that

$$\mathbb{P}(0 < d(W, I) \leq r) < \varepsilon. \tag{3.20}$$

Also,

$$\mathbb{P}(A^{\mathsf{c}} \cap B, d(W, I) > r) \leq \mathbb{P}(d(\langle v, X(t)\rangle e^{-\lambda t}, I) < \eta_t, d(W, I) > r). \tag{3.21}$$

But $\eta_t \to 0$ as $t \to \infty$, so choose t_1 such that $\eta_{t_1} < \frac{r}{2}$ and let $t_0 = \max(t_1, t_{\frac{r}{2}})$. Then, for any $t \geq t_0$, by (3.21),

$$\mathbb{P}(A^{\mathsf{c}} \cap B, d(W, I) > r) < \varepsilon. \tag{3.22}$$

Hence, (3.19), (3.20) and (3.22) imply (3.18). ■
Proof of Claim 2. Notice that the $Z_{i,j}(s_0)$ are mutually independent of $X_i(t)$, and also, for every i, the $Z_{i,j}(s_0)$ are i.i.d. with mean 0 and finite variance. Thus, we can apply the Central Limit Theorem (Theorem 2.8) and assert the claim. ■
This concludes the proof of the lemma. □

Notice that, as a direct consequence of the Lemmas we just proved, we have the following results:

- For a fixed s,
$$\limsup_{t\to\infty} \mathbb{P}\left(W \in I, \frac{Y(t+s)}{\sqrt{\langle v, X(t)\rangle e^{-\lambda t}}} \le y\right) = \limsup_{t\to\infty}(W \in I, A_1 + A_3 \le y), \quad (3.23)$$
and this remains valid when using lim inf in place of lim sup on both sides.
- For every $\eta > 0, \delta > 0$, there exists $s_0(\eta, \delta)$ such that
$$\begin{aligned} \limsup_{t\to\infty} \mathbb{P}(W \in I, A_1 + A_3 \le y) &\le \limsup_{t\to\infty} \mathbb{P}(W \in I, A_1 \le y + \eta) + \delta, \\ \liminf_{t\to\infty} \mathbb{P}(W \in I, A_1 + A_3 \le y) &\ge \liminf_{t\to\infty} \mathbb{P}(W \in I, A_1 \le y - \eta) - \delta \end{aligned} \quad (3.24)$$

Lemma 3.36. *The limit $\sigma(s_0) \to \sigma$ as $s_0 \to \infty$ exists.*

Proof. We have
$$\mathrm{Var}\big(Z_{i,j}(s)e^{-\frac{\lambda s}{2}}\big) = e^{-\lambda s}\mathbb{E}\big[|\langle l_1 \operatorname{Re}(v) + l_2 \operatorname{Im}(v), X(s)\rangle|^2\big] + e^{-\lambda s}\big(l_1 \operatorname{Re}(e^{\lambda_i s}v) + l_2(e^{\lambda_i s}v)\big)^2,$$
where the last term vanishes as $s \to \infty$, since $2\operatorname{Re}(\lambda_i) < \lambda$. The fact that the first term on the right-hand side has a limit is a consequence of a more general result presented in section 4 of [Ath69b]. We omit the details. □

Now, we are ready to prove the theorem:

Proof of Theorem 3.31. Let $\varepsilon > 0$, we fix y and choose η such that
$$\Phi\left(\frac{y+\eta}{\sigma}\right) - \Phi\left(\frac{y-\eta}{\sigma}\right) < \frac{\varepsilon}{2}.$$
Now, we pick s_1 such that, for every $s \ge s_1$,
$$\left|\Phi\left(\frac{y+r\eta}{\sigma(s)}\right) - \Phi\left(\frac{y+r\eta}{\sigma}\right)\right| < \frac{\varepsilon}{2}, \quad \text{for } r = \pm 1.$$
Let $\delta = \frac{\varepsilon}{2}$, $s^* = \max(s_0, s_1)$, where s_0 is as in Lemma 3.33. Then, by 3.23, 3.24 and Lemma 3.35, we get
$$\begin{aligned} \liminf_{t\to\infty} \mathbb{P}\left(W \in I, \frac{Y(t)}{\sqrt{\langle v, X(t)\rangle}} \le y\right) &\le \Phi\left(\frac{y+\eta}{\sigma(s^*)}\right)\mathbb{P}(W \in I) + \frac{\varepsilon}{2}, \\ \limsup_{t\to\infty} \mathbb{P}\left(W \in I, \frac{Y(t)}{\sqrt{\langle v, X(t)\rangle}} \le y\right) &\ge \Phi\left(\frac{y-\eta}{\sigma(s^*)}\right)\mathbb{P}(W \in I) - \frac{\varepsilon}{2}. \end{aligned}$$
Since ε is arbitrary, letting $\varepsilon \to 0$, by our choice of η, we see that $\eta \to 0$ and Theorem 3.31 follows. □

As an important consequence, we get the following corollary.

Corollary 3.37. *Under the conditions of Theorem 3.31,*

$$\mathbb{P}\Big(0 < x_1 \leq W \leq x_2 < \infty, Y(t)e^{-\frac{\lambda t}{2}} \leq x\Big) \xrightarrow{t\to\infty} \int_{x_1}^{x_2} \Phi\left(\frac{x}{\sigma\sqrt{y}}\right) \mathrm{d}_y\, \mathbb{P}(W \leq y).$$

Proof. We condition on $W = y$, which may also have probability 0. By Theorem 3.31, we have

$$\begin{aligned}
\lim_{t\to\infty} \int_{x_1}^{x_2} \mathbb{P}\left(\frac{Y(t)}{\sqrt{\langle X(t), v\rangle}} \leq x \middle| W = y\right) \mathrm{d}_y\, \mathbb{P}(W \leq y) = \\
&= \lim_{t\to\infty} \mathbb{P}\left(0 < x_1 \leq W \leq x_2 < \infty, \frac{Y(t)}{\sqrt{\langle X(t), v\rangle}} \leq x\right) \\
&= \mathbb{P}(0 < x_1 \leq W \leq x_2 < \infty)\Phi\left(\frac{x}{\sigma}\right) \\
&= \int_{x_1}^{x_2} \Phi\left(\frac{x}{\sigma}\right) \mathrm{d}_y\, \mathbb{P}(W \leq y).
\end{aligned}$$

Since the integrand is bounded, we can take the limit inside the integral, and since x_1, x_2 are arbitrary, the asymptotic independence suggested by Theorem 3.31 shows that,

$$\mathbb{P}\left(\frac{Y(t)}{\sqrt{\langle X(t), v\rangle}} \leq x \middle| W = y\right) \xrightarrow{t\to\infty} \Phi\left(\frac{x}{\sigma}\right).$$

Now, by the almost sure convergence of Theorem 3.14, on $W = y$ we expect that

$$\lim_{t\to\infty} \mathbb{P}\left(\frac{Y(t)}{\sqrt{e^{\lambda t}}} \leq x \middle| W = y\right) = \lim_{t\to\infty} \mathbb{P}\left(\frac{Y(t)}{\sqrt{\langle X(t), v\rangle}} \leq \frac{x}{\sqrt{y}} \middle| W = y\right) = \Phi\left(\frac{x}{\sigma\sqrt{y}}\right).$$

Therefore,

$$\begin{aligned}
\lim_{t\to\infty} \mathbb{P}\left(0 < x_1 \leq W \leq x_2 < \infty, \frac{Y(t)}{e^{\frac{\lambda t}{2}}} \leq x\right) &= \lim_{t\to\infty} \int_{x_1}^{x_2} \mathbb{P}\left(\frac{Y(t)}{e^{\frac{\lambda t}{2}}} \leq x \middle| W = y\right) \mathrm{d}_y\, \mathbb{P}(W \leq y) \\
&= \int_{x_1}^{x_2} \Phi\left(\frac{x}{\sigma\sqrt{y}}\right) \mathrm{d}_y\, \mathbb{P}(W \leq y).
\end{aligned}$$

□

Remark 3.38. The notation of the integral taken here corresponds to the *Lebesgue-Stieltjes integral* (also cf. section 3.5 of [LG22]). To write this in a nicer way, notice

that, writing $F_W(y) = \mathbb{P}(W \leq y)$, and writing for a random variable $U \sim \mathcal{N}(o, \sigma^2)$,

$$\begin{aligned}\int_{x_1}^{x_2} \Phi\left(\frac{x}{\sigma\sqrt{y}}\right) \mathrm{d}F_W(y) &= \mathbb{E}\left[\mathbb{1}_{x_1 \leq W \leq x_2} \Phi\left(\frac{x}{\sigma\sqrt{y}}\right)\right] \\ &= \mathbb{E}\left[\mathbb{1}_{x_1 \leq W \leq x_2} \mathbb{P}\left(\sqrt{W}U \leq x \middle| W\right)\right] \\ &= \mathbb{E}\left[\mathbb{1}_{x_1 \leq W \leq x_2} \mathbb{E}\left[\mathbb{1}_{-\infty < \sqrt{W}U < \infty} \middle| W\right]\right] \\ &= \mathbb{E}\left[\mathbb{E}\left[\mathbb{1}_{x_1 \leq W \leq x_2} \mathbb{1}_{-\infty < \sqrt{W}U < \infty} \middle| W\right]\right] \\ &= \mathbb{E}\left[\mathbb{1}_{x_1 \leq W \leq x_2} \mathbb{1}_{-\infty < \sqrt{W}U < \infty}\right] \\ &= \mathbb{P}\big(x_1 \leq W \leq x_2, \sqrt{W}U \leq x\big),\end{aligned}$$

i.e. given that W is an almost surely finite random variable, as stated in Theorem 3.14,

$$e^{-\frac{\lambda t}{2}} \pi_i(X(t)) \xrightarrow{d} \sqrt{W}U.$$

Remark 3.38 implies that the sum of projections P_i, defined earlier, also converges as $t \to \infty$, that is

$$e^{-\frac{\lambda t}{2}} \sum_{i=2}^{d} P_i(X(t)) \xrightarrow{d} \sqrt{W}U, \tag{3.25}$$

where $U \sim \mathcal{N}(0, \Sigma)$. However, computing the covariance matrix Σ is a tough task and may be done using a Martingale central limit theorem (see [Duf13], section 2.1.3 for a reference). Janson presents an even more general result than ours, establishing the convergence with covariance matrix given by

$$\Sigma = \int_0^\infty \sum_{i=2}^{d} P_i^T e^{sR^T} B e^{sR} \sum_{j=2}^{d} P_j e^{-\lambda s} \, \mathrm{d}s \,, \tag{3.26}$$

where B is the $d \times d$ matrix given by $B = \sum_{i=1}^{d} v_{1,i} R_i^T R_i$.

Remark 3.39. In the original paper [Jan04], B is defined considering a d-dimensional *vector of activities* a, so that

$$B = v_{1,i} a_i R_i^T R_i.$$

At the moment, we are not interested in incorporating this vector because, in our case, this vector is $a = (1, \ldots, 1)$, corresponding to our model, in which the balls are drawn uniformly at random. Therefore, we can disregard this component for now, as it does not affect the result, though it will be useful in the proof of Theorem 3.24.

The reason we don't go into detail on this calculation is that, Janson uses a sort of Martingale central limit theorem defining the quadratic variation of a vector-valued process as matrix-valued, which does not align with the method we have used up to this point (see e.g. the proof of Lemma 3.20). What we can do, however, is to simplify this

result using the assumption that R is diagonalizable. So,

$$\begin{aligned}\Sigma &= \int_0^\infty \sum_{j=2}^d e^{s\lambda_j} v_j^T u_j B \sum_{k=2}^d e^{s\lambda_k} u_k^T v_k e^{-\lambda s}\,\mathrm{d}s \\ &= \sum_{j,k=2}^d v_j^T u_j B u_k^T v_k \int_0^\infty e^{s(\lambda_j+\lambda_k-\lambda)}\,\mathrm{d}s\,.\end{aligned}$$

Now, the integral exists, since $\lambda_j + \lambda_k < \frac{\lambda}{2} + \frac{\lambda}{2} = \lambda$, by assumption on the eigenvalues, and we use the fact that we can write,

$$B = \sum_{i=1}^d v_{1,i} R_i^T R_i = R^T D R, \tag{3.27}$$

where $D = \mathrm{diag}(v_{1,1}, v_{1,2}, \ldots, v_{1,d})$. Indeed, for every j,k-th component, we have

$$\sum_{i=1}^d v_{1,i} \big(R_i^T R_i\big)_{j,k} = \sum_{i=1}^d v_{1,i} R_{j,i} R_{k,i} = \big(R^T D R\big)_{j,k}.$$

So,

$$\Sigma = \sum_{j,k=2}^d \frac{u_j R^T D R u_k^T}{\lambda - \lambda_j - \lambda_k} v_j^T v_k = \sum_{j,k=2}^d \frac{\lambda_j \lambda_k u_j D u_k^T}{\lambda - \lambda_j - \lambda_k} v_j^T v_k.$$

Lemma 3.40. *Let $U \sim \mathcal{N}(0, \Sigma)$ as in (3.25). We have,*

$$\frac{1}{\sqrt{z}} \sum_{i=2}^d P_i(X(\tau_b(z))) \xrightarrow{d} \frac{1}{\sqrt{\langle b, v\rangle}} U,$$

where $b \in \mathbb{R}^d$, and $\tau_b(z)$ is the stopping time defined in (3.17).

Proof. As $z \to \infty$, $\tau_b(z) \to \infty$, so let $t = \tau_b(z) - \frac{1}{\lambda}\log(z) + \frac{1}{\lambda}\log(z)$. By Lemma 3.26,

$$\tau_b(z) - \frac{1}{\lambda}\log(z) \xrightarrow{z\to\infty} -\frac{1}{\lambda}(\log(W) + \log(\langle b, v\rangle)) \quad \text{a.s.}$$

So we have that, as $z \to \infty$,

$$\begin{aligned} e^{-\frac{\lambda t}{2}} \sum_{i=2}^d P_i(X(t)) &= e^{-\frac{\lambda}{2}(\tau_b(z) - \frac{1}{\lambda}\log(z) + \frac{1}{\lambda}\log(z))} \sum_{i=2}^d P_i(X(\tau_b(z))) \\ &\approx e^{-\frac{\log(z)}{2}} e^{-\frac{\lambda}{2}(-\frac{1}{\lambda}\log(W) - \frac{1}{\lambda}\log(\langle b,v\rangle))} \sum_{i=2}^d P_i(X(\tau_b(z))) \\ &= \frac{1}{\sqrt{z}} e^{\frac{\log(W)}{2} + \frac{\log(\langle b,v\rangle)}{2}} \sum_{i=2}^d P_i(X(\tau_b(z))). \end{aligned}$$

Thus, using (3.25), we conclude that, as $z \to \infty$,

$$\frac{1}{\sqrt{z}} \sum_{i=2}^{d} P_i(X(\tau_b(z))) \xrightarrow{d} \frac{1}{\sqrt{W}} \cdot \frac{1}{\sqrt{\langle b, v \rangle}} \sqrt{W} U = \frac{1}{\sqrt{\langle b, v \rangle}} U,$$

where $U \sim \mathcal{N}(0, \Sigma)$ as in (3.25). □

Proposition 3.41. *Let $b \in \mathbb{R}^d$ with $\langle b, v \rangle > 0$. Then, as $z \to \infty$,*

$$\frac{1}{\sqrt{z}} \left(X(\tau_b(z)) - \frac{z}{\langle b, v \rangle} v \right) \xrightarrow{d} \mathcal{N}(0, \Sigma_b),$$

where

$$\Sigma_b = \frac{1}{\langle b, v \rangle} \left(\mathrm{Id} - \frac{v^T b}{\langle b, v \rangle} \right) \Sigma \left(\mathrm{Id} - \frac{b^T v}{\langle b, v \rangle} \right). \tag{3.28}$$

Proof. By Proposition 3.30(ii), we have that

$$\frac{1}{\sqrt{z}} \left(X(\tau_b(z) - \frac{z}{\langle b, v \rangle} v \right) = \frac{1}{\sqrt{z}} \sum_{i=2}^{d} P_i(X(\tau_b(z))) \left(\mathrm{Id} - \frac{b^T v}{\langle b, v \rangle} \right) + R(z),$$

for some $R(z) \xrightarrow{z \to \infty} 0$ almost surely. Then, by Lemma 3.40, this converges in distribution to $\frac{1}{\sqrt{z}} \left(\mathrm{Id} - \frac{b^T v}{\langle b, v \rangle} \right) U$, with $U \sim \mathcal{N}(0, \Sigma)$ as in (3.25), where in this case the covariance matrix is given by,

$$\Sigma_b = \frac{1}{\sqrt{\langle b, v \rangle}} \left(\mathrm{Id} - \frac{b^T v}{\langle b, v \rangle} \right)^T \Sigma \frac{1}{\sqrt{\langle b, v \rangle}} \left(\mathrm{Id} - \frac{b^T v}{\langle b, v \rangle} \right),$$

which results in (3.28). □

Now, we can prove the first part of Theorem 3.24.

Theorem 3.42. *As $n \to \infty$,*

$$\frac{U(n) - n\lambda v}{\sqrt{n}} \xrightarrow{d} \mathcal{N}(0, \Sigma_1),$$

where

$$\Sigma_1 = \sum_{j,k=2}^{d} \frac{\lambda u_j D u_k^T}{\lambda - \lambda_j - \lambda_k} w_j^T w_k,$$

for $w_i = \lambda_i v_i - \lambda \|v_i\|_1 v$ and $D = diag(v_{1,1}, v_{1,2}, \ldots, v_{1,d})$.

In order to prove the theorem, we modify the rules of the process by introducing an additional *dummy ball* of a new colour $d+1$ each time a ball of colour $i = 1, \ldots, d$ is drawn/splits. To describe the splitting behaviour, we define the *activity vector* $a = (a_1, \ldots, a_d)$ (mentioned earlier in Remark 3.39), where each component a_i represents the splitting rate of balls of colour i. In our setup however, since balls are drawn uniformly

at random, the activity vector was up until now $a = (1, 1, \dots, 1)$, as all balls have the same probability to be picked. Now, the dummy balls introduced have $\tilde{a}_{d+1} = 0$, where $\tilde{a} = (a, 0)$, meaning that they are never drawn (if this is the case, one can simply place the ball back into the urn, pretend the event never occurred, and repeat the drawing for that time step). Also, we modify our replacement matrix such that $\tilde{R}_{d+1} = 0$ and $\tilde{R}^T_{d+1} = \tilde{a}^T$, i.e.

$$\tilde{R} = \begin{pmatrix} R & a^T \\ 0 & 0 \end{pmatrix}.$$

This adjustment does not affect the behaviour of balls of colour $i = 1, \dots, d$, but it allows to keep track of the number of splits in the system. Notice that, now the process is not irreducible, however the non-extinction (tenability) assumption still hold. For the modified process, we write $\tilde{}$ over the inherent symbols. Notice that, by definition of $\tilde{R}$, we have $\tilde{R}_i = (R_i, 1)$ for every $i = 1, \dots, d$. The set of eigenvalues has now also the value $\{0\}$, in particular $\tilde{\lambda} = \lambda$. It's easy to see that, the associated eigenvectors are $\tilde{v} = (v, \frac{1}{\lambda})$ and $\tilde{u} = (u_1, 0)$.
For later purposes, we define the following (matrix-valued) functions:

$$f(s, R) = \sum_{n=1}^{\infty} \frac{s^n}{n!} R^{n-1} = \int_0^s e^{tR} \, \mathrm{d}t \quad \text{and} \quad g(s, R) = e^{sR} - \lambda f(s, R) a^T v.$$

Lemma 3.43. *We have the following properties for $f(s, R)$ and $g(s, R)$,*

(i) $Rf(s, R) = e^{sR} - 1$,

(ii) $Rg(s, R) = e^{sR}(R - \lambda a^T v) + \lambda a^T v$

(iii) $vg(s, R) = v$

Proof. Part (i) follows by definition. Part (ii) follows similarly by definition and using part (i). For part (iii), notice that, by definition

$$vf(s, R) = \int_0^s ve^{tR} \, \mathrm{d}t = \int_0^s e^{t\lambda} v \, \mathrm{d}t = \frac{1}{\lambda} v(e^{s\lambda} - 1).$$

Hence,

$$vg(s, R) = e^{\lambda s} v - (e^{\lambda s} - 1) v a^T v = v.$$

□

Proof of Theorem 3.42. We want to apply Proposition 3.41 to the modified process, because notice that, choosing $b = (0, \dots, 0, 1) \in \mathbb{R}^{d+1}$, we get that $\tau_b(n)$ is the first time we have n dummy balls. In other words, this is the n-th split time τ_n, and thus $\tilde{U}(n) = \tilde{X}(\tau_b(n))$. Moreover, we get that $\langle b, \tilde{v} \rangle = \frac{1}{\lambda}$. Therefore, by Proposition 3.41, we get a convergence to a Gaussian limit $\tilde{V}$. Ignoring the dummy balls, the limit is given by $V = \tilde{V}(\mathrm{Id}, 0)^T$, where $V \sim \mathcal{N}(0, \Sigma_1)$, i.e.

$$\frac{U(n) - n\lambda v}{\sqrt{n}} \xrightarrow{d} \mathcal{N}(0, \Sigma_1).$$

We proceed now with the computation of Σ_1. First, we write Σ_b in terms of the original expression for Σ, given in (3.26), that is

$$\Sigma_b = \frac{1}{\langle b, v\rangle}\int_0^\infty \left(\mathrm{Id} - \frac{v^T b}{\langle b, v\rangle}\right) e^{sR^T} B e^{sR}\left(\mathrm{Id} - \frac{b^T v}{\langle b, v\rangle}\right) e^{-\lambda s}\,\mathrm{d}s\,, \tag{3.29}$$

which follows by noticing that,

$$\sum_{i=2}^{d} P_i = \mathrm{Id} - P_1 \quad \text{and} \quad \big(\mathrm{Id} - P_1\big)\left(\mathrm{Id} - \frac{b^T v}{\langle b, v\rangle}\right) = \left(\mathrm{Id} - \frac{b^T v}{\langle b, v\rangle}\right).$$

Now,

$$\tilde{R}_i^T \tilde{R}_i = \begin{pmatrix} R_i^T R_i & R_i^T \\ R_i & 1 \end{pmatrix},$$

so,

$$\begin{aligned} \tilde{B} &= \sum_{i=1}^{d+1} \tilde{v}_{1,i}\tilde{a}_i\tilde{B}_i = \sum_{i=1}^{d} v_{1,i}a_i\tilde{B}_i = \sum_{i=1}^{d} v_{1,i}\tilde{B}_i \\ &= \begin{pmatrix} \sum_{i=1}^{d} v_{1,i}R_i^T R_i & \sum_{i=1}^{d} v_{1,i}R_i^T \\ \sum_{i=1}^{d} v_{1,i}R_i & \|v\|_1 \end{pmatrix} \\ &= \begin{pmatrix} B & \lambda v^T \\ \lambda v & 1 \end{pmatrix}, \end{aligned} \tag{3.30}$$

where we used the fact that

$$\sum_{i=1}^{d} v_{1,i}R_i = \Big(\sum_{i=1}^{d} v_{1,i}R_{i,j}\Big)_{1\le j\le d} = vR = \lambda v.$$

By induction, we have that

$$\tilde{R}^n = \begin{pmatrix} R^n & R^{n-1}a^T \\ 0 & 0 \end{pmatrix} \quad \forall n \ge 1,$$

thus,

$$e^{s\tilde{R}} = \sum_{n=0}^{\infty} \frac{s^n}{n!}\tilde{R}^n = \begin{pmatrix} e^{sR} & f(s,R)a^T \\ 0 & 1 \end{pmatrix}.$$

Now,

$$\begin{aligned} e^{s\tilde{R}}\left(\mathrm{Id} - \frac{b^T\tilde{v}}{\langle b, \tilde{v}\rangle}\right) &= e^{s\tilde{R}}\left(\begin{pmatrix} \mathrm{Id} & 0 \\ 0 & 1 \end{pmatrix} - \lambda\begin{pmatrix} 0 \\ 1 \end{pmatrix}\begin{pmatrix} v & \frac{1}{\lambda} \end{pmatrix}\right) \\ &= \begin{pmatrix} e^{sR} & f(s,R)a^T \\ 0 & 1 \end{pmatrix}\begin{pmatrix} \mathrm{Id} & 0 \\ -\lambda v & 0 \end{pmatrix} \\ &= \begin{pmatrix} g(s,R) & 0 \\ -\lambda v & 0 \end{pmatrix}. \end{aligned} \tag{3.31}$$

So, by relation (3.29),

$$\begin{aligned}\Sigma_1 &= \lambda \int_0^\infty \begin{pmatrix} g(s,R)^T & -\lambda v^T \\ 0 & 0 \end{pmatrix} \begin{pmatrix} B & \lambda v^T \\ \lambda v & 1 \end{pmatrix} \begin{pmatrix} g(s,R) & 0 \\ -\lambda v & 0 \end{pmatrix} e^{-\lambda s}\,\mathrm{d}s \\ &= \lambda \int_0^\infty \big(g(s,R)^T B g(s,R) - \lambda^2 v^T v g(s,R) - \lambda^2 g(s,R)^T v^T v + \lambda^2 v^T v \big) e^{-\lambda s}\,\mathrm{d}s \\ &= \lambda \int_0^\infty \big(g(s,R)^T B g(s,R) - \lambda^2 v^T v \big) e^{-\lambda s}\,\mathrm{d}s \\ &= \lambda \int_0^\infty g(s,R)^T B g(s,R) e^{-\lambda s}\,\mathrm{d}s - \lambda^2 v^T v,\end{aligned}$$

where the third equality follows applying Lemma 3.43(iii). Now, using relation 3.27 and Lemma 3.43(ii),

$$\begin{aligned}g(s,R)^T B g(s,R) = (R-\lambda a^T v)^T e^{sR^T} D e^{sR} (R - \lambda a^T v) + \lambda (R-\lambda a^T v)^T e^{sR^T} D a^T v \\ + \lambda v^T a D e^{sR} (R - \lambda a^T v) + \lambda^2 v^T a D a^T v.\end{aligned} \tag{3.32}$$

It is clear that $aD = v$, thus $aDa^T = \|v\|_1 = 1$, and $aDR = vR = \lambda v$. It follows that, $(aD - v)e^{sR}(R - \lambda a^T v) = 0$, or, in other terms,

$$a D e^{sR} (R - \lambda a^T v) = v e^{sR} (R - \lambda a^T v) = 0,$$

where the last equality is a consequence of $ve^{sR} = e^{s\lambda} v$, $vR = \lambda v$ and $va^T = \|v\| = 1$. So, the second and third terms in the sum of the right-hand side of (3.32) vanish, thus we are left with

$$g(s,R)^T B g(s,R) = (R-\lambda a^T v)^T e^{sR^T} D e^{sR} (R - \lambda a^T v) + \lambda^2 v^T v.$$

Moreover,

$$e^{sR}(R - \lambda a^T v) = \sum_{i=1}^d u_i^T v_i e^{sR}(R - \lambda a^T v) = \sum_{i=2}^d e^{s\lambda_i} u_i^T (\lambda_i v_i - \lambda \|v_i\|_1 v) =: \sum_{i=2}^d e^{s\lambda_i} u_i^T w_i.$$

Therefore,

$$\begin{aligned}\Sigma_1 &= \lambda \int_0^\infty g(s,R)^T B g(s,R) e^{-\lambda s}\,\mathrm{d}s - \lambda^2 v^T v \\ &= \lambda \int_0^\infty \sum_{i=2}^d \sum_{j=2}^d e^{s\lambda_i} e^{s\lambda_j} w_i^T u_i D u_j^T w_j e^{-\lambda s}\,\mathrm{d}s \\ &= \sum_{i,j=2}^d \frac{\lambda u_i D u_j^T}{\lambda - \lambda_i - \lambda_j} w_i^T w_j.\end{aligned}$$

□

This ends the proof of Theorem 3.24(i), and thus the section about small eigenvalues. We continue now, with the similar section about the case $2\,\mathrm{Re}(\lambda_i) = \lambda$.

3.3.4.2 The case $\mathrm{Re}(\lambda_i) = \frac{\lambda}{2}$

This case is very similar to the previous case, regarding small eigenvalues. Throughout this section we assume that $\mathrm{Re}(\lambda_i) = \lambda/2$ for $i = 2, \dots, d$. As we will shortly see, in this case we also need to consider a logarithmic factor. For brevity, we omit the details (one can confront the full proof on [AN72], ch.V, Theorem 8.4). For this case, we limit ourselves to presenting the results in a manner similar to the previous section, up to the expression regarding the stopping time $\tau_b(z)$, and to proving only Theorem 3.24.

Theorem 3.44. *Let $A = \{0 < x_1 \leq W \leq x_2 < \infty\}$, under the hypothesis of Theorem 3.31,*

$$\mathbb{P}\left(A, \frac{Y(t)}{\sqrt{\langle X(t), v\rangle \log(\langle X(t), v\rangle)}} \leq x\right) \xrightarrow{t\to\infty} \mathbb{P}(A)\Phi\left(\frac{x}{\sigma}\right).$$

Reasoning exactly as in Corollary 3.37 and Remark 3.38 leads us to the following.

Corollary 3.45. *As $t \to \infty$,*

$$t^{-\frac{1}{2}} e^{-\frac{\lambda t}{2}} \pi_i(X(t)) \xrightarrow{d} \sqrt{W}\mathcal{N}(0, \sigma^2).$$

Hence, similarly in [Jan04], if $\mathrm{Re}(\lambda_i) = \frac{\lambda}{2}$, for any $i = 2, \dots, d$,

$$t^{-\frac{1}{2}} e^{-\lambda_i t} P_i(X(t)) \xrightarrow{d} \sqrt{W} V_i, \tag{3.33}$$

as $t \to \infty$, where $V_i \sim \mathcal{N}(0, \Sigma')$ and $\Sigma' = \sum_{i=2}^{d} P_i^* B P_i$. Here P^* denotes the Hermitian conjugate of P.

Lemma 3.46. *Let $b \in \mathbb{R}^d$ with $\langle b, v\rangle > 0$. Then, as $z \to \infty$,*

$$\big(z\log(z)\big)^{-\frac{1}{2}} P_i(X(\tau_b(z))) \xrightarrow{d} \frac{1}{\sqrt{\langle b, v\rangle}} V_i',$$

where $V_i' = \frac{1}{\sqrt{\lambda}} V_i$.

Proof. Notice that, writing $t = \frac{\tau_b(z)\lambda}{\log(z)} \cdot \frac{\log(z)}{\lambda} - \frac{\log(z)}{\lambda} + \frac{\log(z)}{\lambda}$, by Lemma 3.26,

$$\frac{\tau_b(z)\lambda}{\log(z)} \xrightarrow{z\to\infty} 1 \quad \text{a.s. and} \quad \tau_b(z) - \frac{\log(z)}{\lambda} \xrightarrow{z\to\infty} -\frac{1}{\lambda}(\log(W) + \log(\langle b, v\rangle)) \quad \text{a.s.}$$

The results follows by (3.33), analogously as we did in the proof of Lemma 3.40. □

Proposition 3.47. *Under the assumptions of Lemma 3.46,*

$$\frac{1}{\sqrt{z\log(z)}}\left(X(\tau_b(z)) - \frac{z}{\langle b, v\rangle} v\right) \xrightarrow{d} \mathcal{N}(0, \Sigma_b'),$$

where

$$\Sigma_b' = \frac{1}{\lambda\langle b, v\rangle}\left(\mathrm{Id} - \frac{v^T b}{\langle b, v\rangle}\right)\Sigma'\left(\mathrm{Id} - \frac{b^T v}{\langle b, v\rangle}\right).$$

Proof. As in the proof of Proposition 3.41, the result follows applying Proposition 3.30(ii) and Lemma 3.46. The application results in the Gaussian limit $\frac{1}{\sqrt{\langle b,v\rangle}}V_i'$, which yields our result. □

Finally, we can prove the main result of this case.

Theorem 3.48. *As $n \to \infty$,*

$$\frac{U(n) - n\lambda v}{\sqrt{n\log(n)}} \xrightarrow{d} \mathcal{N}(0, \Sigma_2),$$

where

$$\Sigma_2 = \sum_{i=2}^{d} \bar{u}_i D u_i^T w_i^* w_i,$$

for $w_i = \lambda_i v_i - \lambda \|v_i\|_1 v$, and $\bar{u}$ denotes the complex conjugate of a vector u.

Proof. Arguing as in the proof of Theorem 3.42, now applying Proposition 3.47, we have $\mathrm{Re}(\tilde{\lambda}_2) = \mathrm{Re}(\lambda_2) = \lambda/2$. So,

$$\frac{U(n) - n\lambda v}{\sqrt{n\log(n)}} \xrightarrow{d} V = \tilde{V}_b(\mathrm{Id}, 0)^T,$$

where $\tilde{V}_b$ is Gaussian.

Claim.

$$\frac{1}{t}\int_0^t \big(\mathrm{Id} - P_i\big)^* e^{sR^T} B e^{sR} \big(\mathrm{Id} - P_i\big) e^{-\lambda s}\,\mathrm{d}s \xrightarrow{t\to\infty} \Sigma'.$$

Proof. Indeed, the left-hand side is equal to

$$\sum_{i=2}^{d} \frac{1}{t}\int_0^t P_i^* e^{s\bar{\lambda}_i} B e^{s\lambda_i} P_i e^{-s\lambda}\,\mathrm{d}s = \sum_{i=2}^{d} \frac{1}{t}\int_0^t e^{(\lambda_i + \bar{\lambda}_i - \lambda)s}\,\mathrm{d}s\, P_i^* B P_i.$$

The claim follows by noticing that $\lambda_i + \bar{\lambda}_i - \lambda = 2\,\mathrm{Re}(\lambda_i) - \lambda = 0$. ■

The claim implies, by multiplying to the left by $\big(\mathrm{Id} - \frac{v^T b}{\langle b,v\rangle}\big)$ and to the right by its transpose, and since $P_1\big(\mathrm{Id} - \frac{b^T v}{\langle b,v\rangle}\big) = 0$, that

$$\frac{1}{t}\int_0^t \left(\mathrm{Id} - \frac{v^T b}{\langle b,v\rangle}\right) e^{sR^T} B e^{sR} \left(\mathrm{Id} - \frac{b^T v}{\langle b,v\rangle}\right) e^{-\lambda s}\,\mathrm{d}s \xrightarrow{t\to\infty} \left(\mathrm{Id} - \frac{v^T b}{\langle b,v\rangle}\right)\Sigma' \left(\mathrm{Id} - \frac{b^T v}{\langle b,v\rangle}\right).$$

So, by Proposition 3.47 and using the claim,

$$\begin{aligned}\tilde{\Sigma}_b' &= \left(\mathrm{Id} - \frac{\tilde{v}^T b}{\langle b,\tilde{v}\rangle}\right)\tilde{\Sigma}'\left(\mathrm{Id} - \frac{b^T \tilde{v}}{\langle b,\tilde{v}\rangle}\right) \\ &= \lim_{t\to\infty}\frac{1}{t}\int_0^t \left(\mathrm{Id} - \frac{\tilde{v}^T b}{\langle b,\tilde{v}\rangle}\right) e^{s\tilde{R}^T} \tilde{B} e^{s\tilde{R}} \left(\mathrm{Id} - \frac{b^T \tilde{v}}{\langle b,\tilde{v}\rangle}\right) e^{-\lambda s}\,\mathrm{d}s\,.\end{aligned}$$

Hence, by equations (3.30), (3.31) and Lemma 3.43(iii),

$$\begin{aligned}\Sigma_2 &= \lim_{t\to\infty}\frac{1}{t}\int_0^t \begin{pmatrix} g(s,R)^T & -\lambda v^T\end{pmatrix}\begin{pmatrix} B & \lambda v^T \\ \lambda v & 1\end{pmatrix}\begin{pmatrix} g(s,R) & -\lambda v\end{pmatrix}^T e^{-\lambda s}\,\mathrm{d}s \\ &= \lim_{t\to\infty}\frac{1}{t}\int_0^t \big(g(s,R)^T B g(s,R) - \lambda^2 v^T v\big)e^{-\lambda s}\,\mathrm{d}s\,.\end{aligned}$$

Writing $g(s,R)=\sum_{i=1}^d P_i g(s,R)$ and since $2\,\mathrm{Re}(\lambda_i)=\lambda$, we get

$$\begin{aligned}\Sigma_2 = \lim_{t\to\infty}\frac{1}{t}\int_0^t \Bigg(\sum_{i=1}^d \Big(e^{s\bar{\lambda_i}}P_i^* - \frac{\lambda}{\bar{\lambda_i}}\Big(e^{s\bar{\lambda_i}}-1\Big)v^T a P_i^*\Big)B\Big(e^{s\lambda_i}P_i - \frac{\lambda}{\lambda_i}\Big(e^{s\lambda_i}-1\Big)P_i a^T v\Big) \\ - \lambda^2 v^T v\Bigg)e^{-\lambda s}\,\mathrm{d}s \\ = \lim_{t\to\infty}\frac{1}{t}\int_0^t \Bigg(\Bigg(\sum_{i=2}^d e^{s\bar{\lambda_i}}\Big(P_i^* - \frac{\lambda}{\bar{\lambda_i}}v^T a P_i^* + e^{-s\bar{\lambda_i}}\frac{\lambda}{\bar{\lambda_i}}v^T a P_i^*\Big)B e^{s\lambda_i}\Big(P_i - \frac{\lambda}{\lambda_i}P_i a^T v\Big) \\ + e^{-s\lambda_i}\frac{\lambda}{\lambda_i}P_i a^T v\Bigg)e^{-\lambda s} - \lambda^2 v^T v e^{-\lambda s}\Bigg)\,\mathrm{d}s\,.\end{aligned}$$

Now, since $e^{s(\bar{\lambda_i}+\lambda_i-\lambda)}=1$, and noticing that, the limit of the integral of the terms involving $e^{-s\bar{\lambda_i}}, e^{-s\lambda_i}$, and the last term $\lambda^2 v^T v e^{-\lambda s}$, go to zero, we have,

$$\begin{aligned}\Sigma_2 &= \sum_{i=2}^d \Big(P_i^* - \frac{\lambda}{\bar{\lambda_i}}v^T a P_i^*\Big)B\Big(P_i - \frac{\lambda}{\lambda_i}P_i a^T v\Big) \\ &= \sum_{i=2}^d \Big(\mathrm{Id} - \frac{\lambda}{\bar{\lambda_i}}v^T a\Big)P_i^* B P_i\Big(\mathrm{Id} - \frac{\lambda}{\lambda_i}a^T v\Big).\end{aligned}$$

Then, since $P_i = u_i^T v_i$, and $B = R^T D R$, as before,

$$\begin{aligned}\Sigma_2 &= \sum_{i=2}^d \big(\bar{\lambda_i}v_i^* - \lambda\langle a, v_i\rangle v^T\big)\bar{u_i}D u_i^T\big(\lambda_i v_i - \lambda\langle a, v_i\rangle v\big) \\ &=: \sum_{i=2}^d \bar{u_i}D u_i^T w_i^* w_i.\end{aligned}$$

□

This ends the proof of Theorem 3.24(ii), and thus this section. Now we can finish with the section, dedicated to big eigenvalues.

3.3.4.3 Big eigenvalues

Throughout this section we assume that $\mathrm{Re}(\lambda_i) > \lambda/2$ for $i = 2,\dots,d$. Recall that we have already proved some results for this case in section 3.3.3, these are Lemma 3.20 and Lemma 3.21. In particular, as a remark, we have that:

Remark 3.49. Lemmas 3.20 and 3.21 hold also for the projections P_i, that is

$$e^{-\lambda_i t} P_i(X(t)) \xrightarrow{t\to\infty} W v_i \quad \text{a.s.}$$

Lemma 3.50. *As $z \to \infty$, almost surely,*

$$z^{-\frac{\lambda_i}{\lambda}} P_i(X(\tau_b(z))) \to \langle b, v\rangle^{-\frac{\lambda_i}{\lambda}} W^{1-\frac{\lambda_i}{\lambda}} v_i.$$

Proof. Similarly as before, we write $t = \tau_b(z) - \frac{1}{\lambda}\log(z) + \frac{1}{\lambda}\log(z)$, and recall that, by Lemma 3.26,

$$\tau_b(z) - \frac{1}{\lambda}\log(z) \xrightarrow{z\to\infty} -\frac{1}{\lambda}(\log(W) + \log(\langle b, v\rangle)) \quad \text{a.s.}$$

So, as $z \to \infty$,

$$\begin{aligned} e^{-\lambda_i t} P_i(X(t)) &= e^{-\lambda_i(\tau_b(z) - \frac{1}{\lambda}\log(z) + \frac{1}{\lambda}\log(z))} P_i(X(\tau_b(z))) \\ &\approx e^{-\frac{\lambda_i}{\lambda}(\log(W) + \log(\langle b,v\rangle)) - \frac{\lambda_i}{\lambda}\log(z)} P_i(X(\tau_b(z))) \\ &= z^{-\frac{\lambda_i}{\lambda}} \langle b, v\rangle^{-\frac{\lambda_i}{\lambda}} W^{-\frac{\lambda_i}{\lambda}}. \end{aligned}$$

Thus, by Lemma 3.20 and Remark 3.49,

$$z^{-\frac{\lambda_i}{\lambda}} P_i(X(\tau_b(z))) \xrightarrow{z\to\infty} \langle b, v\rangle^{-\frac{\lambda_i}{\lambda}} W^{1-\frac{\lambda_i}{\lambda}} v_i.$$

□

In what follows, instead of using i as the index previously employed, we will use j to avoid ambiguity with the imaginary unit $i = \sqrt{-1}$.

Proposition 3.51. *Let $b \in \mathbb{R}^d$ with $\langle b, v\rangle > 0$. Then,*

$$\mathcal{X}_b(z) := z^{-\frac{\mathrm{Re}(\lambda_2)}{\lambda}} \left(X(\tau_b(z)) - \frac{z}{\langle b, v\rangle} v \right) \text{ is tight,}$$

in the sense that, considering $\Lambda := \{\lambda_j : \mathrm{Re}(\lambda_j) = \mathrm{Re}(\lambda_2), \mathrm{Im}(\lambda_j) \geq 0\}$, there exist complex random variables W_{b,λ_j}, such that,

$$\mathcal{X}_b(z) - \mathrm{Re}\left(\sum_{\lambda_j \in \Lambda} e^{i\frac{\mathrm{Im}(\lambda_j)}{\lambda}\log(z)} W_{b,\lambda_j} \right) \xrightarrow{z\to\infty} 0 \quad \text{a.s.} \tag{3.34}$$

In particular, if $\lambda_2 \in \mathbb{R}$ and $\Lambda = \{\lambda_2\}$, $\mathcal{X}_b(z) - W_b \to 0$ a.s. for some random variable W_b, and thus

$$\mathcal{X}_b(z) \xrightarrow{d} W_b,$$

as $z \to \infty$.

Proof. By Proposition 3.30(ii), it follows that

$$\mathcal{X}_b(z) - \sum_{j=2}^{d} z^{-\frac{\mathrm{Re}(\lambda_2)}{\lambda}} P_j(X(\tau_b(z)))\left(\mathrm{Id} - \frac{b^T v}{\langle b, v\rangle}\right) \xrightarrow{z\to\infty} 0 \quad \text{a.s.}$$

By Lemmas 3.40, 3.46, 3.50, the terms in the sum with $\mathrm{Re}(\lambda_j) \neq \mathrm{Re}(\lambda_2)$ tend to 0 in probability. One can see this, noticing that, since we assumed $\mathrm{Re}(\lambda_2)/\lambda > 1/2$, and hence, if $\mathrm{Re}(\lambda_j) < \mathrm{Re}(\lambda_2)$, then the $P_j(X(\tau_b(z)))$, for $j \geq 3$, are of smaller order than $z^{\frac{\mathrm{Re}(\lambda_2)}{\lambda}}$. By Lemma 3.50 and the arguments presented in the proof of Lemma 3.22, this convergence in probability is improved to almost sure convergence. Moreover, by Lemma 3.50, if $\mathrm{Re}(\lambda_j) = \mathrm{Re}(\lambda_2)$,

$$z^{-\frac{\mathrm{Re}(\lambda_2)}{\lambda}} P_j(X(\tau_b(z))) - z^{i\frac{\mathrm{Im}(\lambda_j)}{\lambda}} \langle b, v\rangle^{-\frac{\lambda_j}{\lambda}} W^{1-\frac{\lambda_j}{\lambda}} v_j \xrightarrow{z\to\infty} 0 \quad \text{a.s.}$$

which implies,

$$\mathcal{X}_b(z) - \sum_{\lambda_j\in\Lambda} z^{i\frac{\mathrm{Im}(\lambda_j)}{\lambda}} \hat{W} \xrightarrow{z\to\infty} 0$$

almost surely, where

$$\hat{W} = \langle b, v\rangle^{-\frac{\lambda_j}{\lambda}} W^{1-\frac{\lambda_j}{\lambda}} v_j \left(\mathrm{Id} - \frac{b^T v}{\langle b, v\rangle}\right).$$

Then, (3.34) follows with $W_{b,\lambda_j} = \hat{W}$, if $\mathrm{Im}(\lambda_j) = 0$, and $W_{b,\lambda_j} = 2\hat{W}$, if $\mathrm{Im}(\lambda_j) > 0$. The last assertion of the proposition follows as a special case, when $\Lambda = \{\lambda_2\}$ with $\lambda_2 \in \mathbb{R}$. □

We state now the last result, which concludes this section.

Theorem 3.52. *Under the assumptions of this section,*

$$\mathcal{U}(n) := n^{-\frac{\mathrm{Re}(\lambda_2)}{\lambda}}\big(U(n) - n\lambda v\big) \text{ is tight,}$$

and we have the same type of asymptotic behaviour as in Proposition 3.51. In particular, if $\Lambda = \{\lambda_2\}$, $\mathcal{U}(n) - \mathrm{Re}\left(n^{i\frac{\mathrm{Im}(\lambda_2)}{\lambda}}\hat{W}\right) \to 0$ *almost surely, for some random vector* $\hat{W}$. *Thus, if* $\Lambda = \{\lambda_2\}$ *with* $\lambda_2 \in \mathbb{R}$, *as* $n \to \infty$,

$$\mathcal{U}(n) \xrightarrow{d} \hat{W}.$$

Proof. This follows directly by applying Proposition 3.51 to $\tilde{U}(n)$, analogously as before. □

Remark 3.53. We do not have a general description of the random vector $\hat{W}$ in Theorem 3.52. However, in [Jan04], its moments are computed for specific examples, as $\hat{W}$ is related in some way to the random variable W. We omit these computations here.

This concludes this case and the results of the section about a central limit theorem for the irreducible case. We now present some examples (present in [Jan04]) where these results can be applied.

Example 3.54 (Friedman's urn)**.** Let $U(n)$ be Friedman's urn ([Fri49]), where we consider only two colours, so $d = 2$, with replacement matrix

$$R = \begin{pmatrix} \alpha & \beta \\ \beta & \alpha \end{pmatrix}.$$

We assume $\alpha \geq -1$ and $\beta > 0$, so that R is indeed an irreducible matrix. The eigenvalues and associated eigenvectors are given by

$$\begin{aligned} \lambda_1 = (\alpha + \beta) \quad &\text{and} \quad \lambda_2 = (\alpha - \beta), \\ v_1 = \left(\frac{1}{2}, \frac{1}{2}\right) \quad &\text{and} \quad v_2\left(\frac{1}{2}, -\frac{1}{2}\right), \\ u_1 = \big(1, 1\big) \quad &\text{and} \quad u_2\big(1, -1\big). \end{aligned}$$

Now, if $\lambda_2 < \frac{\lambda_1}{2}$, equivalently $\alpha < 3\beta$, then, by Theorem 3.42, as $n \to \infty$,

$$\frac{U(n) - n(\alpha + \beta)\left(\frac{1}{2}, \frac{1}{2}\right)}{\sqrt{n}} \xrightarrow{d} \mathcal{N}(0, \Sigma_1),$$

where the covariance matrix is given by,

$$\Sigma_1 = \frac{\alpha + \beta}{3\beta - \alpha} \cdot \frac{(\alpha - \beta)^2}{4} \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix}.$$

Next, if $\lambda_2 = \frac{\lambda_1}{2}$, $\alpha = 3\beta$, Theorem 3.48 yields as $n \to \infty$,

$$\frac{U(n) - n(\alpha + \beta)\left(\frac{1}{2}, \frac{1}{2}\right)}{\sqrt{n \log(n)}} \xrightarrow{d} \mathcal{N}(0, \Sigma_2),$$

where

$$\Sigma_2 = \frac{(\alpha - \beta)^2}{4} \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix}.$$

Lastly, if $\lambda_2 > \frac{\lambda}{2}$, $\alpha > 3\beta$, as $n \to \infty$, by Theorem 3.52,

$$n^{-\frac{\alpha - \beta}{\alpha + \beta}} \left(U(n) - n(\alpha + \beta)\left(\frac{1}{2}, \frac{1}{2}\right)\right) \to \hat{W} \quad \text{a.s.}$$

for some random vector $\hat{W}$.
Some of these results were already shown by several authors, e.g. by Bernstein ([Ber40]) and Freedman ([Fre65]).

Example 3.55 (Generalized Friedman's urn)**.** We consider now a generalization of the previous case. As before, $d = 2$ and let $U(n)$ denote the urn with replacement matrix

$$R = \begin{pmatrix} \alpha & \beta \\ \gamma & \zeta \end{pmatrix}.$$

We assume $\gamma, \beta > 0$, so that R is irreducible, and $\alpha, \zeta \geq -1$. To simplify what follows, we write $\kappa = \frac{\alpha+\zeta}{2}, \xi = \frac{\alpha-\zeta}{2}, \rho = \sqrt{\xi^2 + \beta\gamma} > 0$, so we can write the replacement matrix as

$$R = \begin{pmatrix} \kappa + \xi & \beta \\ \gamma & \kappa - \xi \end{pmatrix}.$$

The eigenvalues of R and their associated eigenvectors are then given by,

$$\lambda_{1,2} = \kappa \pm \rho, \quad v_{1,2} = \frac{1}{c_{1,2}}\big(\xi \pm \rho, \beta\big), \quad u_{1,2} = \frac{c_{1,2}}{(\xi \pm \rho)^2 + \beta\gamma}\big(\xi \pm \rho, \gamma\big),$$

for some normalization constants $c_{1,2}$. We choose $c_1 = \beta + \rho + \xi$, so that $\|v_1\|_1 = 1$, and $c_2 = 1$. Then, if $\lambda_2 < \frac{\lambda_1}{2}$, equivalently $\kappa < 3\rho$, Theorem 3.42 yields,

$$\frac{U(n) - n\frac{(\kappa+\rho)}{\xi+\rho+\beta}(\xi + \rho, \beta)}{\sqrt{n}} \xrightarrow{d} \mathcal{N}(0, \Sigma_1),$$

where

$$\Sigma_1 = \frac{(\kappa + \rho)\frac{1}{((\xi-\rho)^2+\beta\gamma)^2}\frac{1}{\xi+\rho+\beta}}{3\rho - \kappa}(\xi - \rho, \gamma)\begin{pmatrix} \xi + \rho & 0 \\ 0 & \beta \end{pmatrix}\begin{pmatrix} \xi - \rho \\ \gamma \end{pmatrix} w_2^T w_2,$$

with $w_2 = (\kappa - \rho)(\xi - \rho, \beta) - \frac{(\kappa+\rho)(\xi-\rho+\beta)}{\xi+\rho+\beta}(\xi + \rho, \beta)$. Using the relations

$$(\rho - \xi)^2 + \beta\gamma = 2\rho(\rho - \xi), \quad (\rho - \xi)(\beta + \rho + \xi) = \beta(\rho + \gamma - \xi), \quad \rho^2 - \xi^2 = \beta\gamma,$$

a simple computation gives us

$$u_2 D u_2^T = \frac{\gamma}{4\rho^2(\rho - \xi)} \quad \text{and} \quad w_2 = \frac{2\rho\beta}{\beta + \rho + \xi}(\gamma - \alpha, \zeta - \beta).$$

Hence, we have

$$\Sigma_1 = \frac{(\kappa + \rho)\beta\gamma}{(3\rho - \kappa)(\beta + \rho + \xi)(\gamma + \rho - \xi)}\begin{pmatrix} (\gamma - \alpha)^2 & (\gamma - \alpha)(\zeta - \beta) \\ (\gamma - \alpha)(\zeta - \beta) & (\zeta - \beta)^2 \end{pmatrix}.$$

Now, if $\lambda_2 = \frac{\lambda_1}{2}$, $\kappa = 3\rho$, then, by Theorem 3.48,

$$\frac{U(n) - n\frac{(\kappa+\rho)}{\xi+\rho+\beta}(\xi + \rho, \beta)}{\sqrt{n\log(n)}} \xrightarrow{d} \mathcal{N}(0, \Sigma_2),$$

where, by similar computations done for the previous case, and the relations above,

$$\Sigma_2 = \frac{\beta\gamma}{(\beta + \rho + \xi)(\gamma + \rho - \xi)}\begin{pmatrix} (\gamma - \alpha)^2 & (\gamma - \alpha)(\zeta - \beta) \\ (\gamma - \alpha)(\zeta - \beta) & (\zeta - \beta)^2 \end{pmatrix}.$$

In the special case, where we assume $\alpha + \beta = \gamma + \zeta = c$, we have $\rho = \frac{\beta+\gamma}{2}$, so $\lambda_1 = \rho + \kappa = c$ and $3\rho - \kappa = c + 2\rho - 2\kappa = c + 2(\gamma - \alpha)$, thus

$$\Sigma_1 = \frac{c\beta\gamma}{(c + 2(\gamma - \alpha))(\beta + \gamma)^2}(\gamma - \alpha)^2 \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix},$$

as well as, when $\alpha = \beta + 2\gamma, \zeta = 2\beta + \gamma$,

$$\Sigma_2 = \beta\gamma \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix},$$

in accordance with [BP85].

Chapter 4

Infinitely many colour Pólya urns

In this chapter, we will address the case analogous to the previous one, but this time with an infinite number of colours. We will refer to these infinitely many colour Pólya urns as Measure-Valued Pólya Processes (MVPPs for short). First we recall what a *Polish space* is ([Sri08], pp.52-57): a Polish space is a complete metric space that is separable, meaning that it has a countable dense subset. Polish spaces are a convenient setting for more advanced measure theory, in particular in probability theory.

Definition 4.1 (**MVPP**)**.** Let E be a Polish space, m_0 be a measure such that $m_0(E) < \infty$ and let $R^{(1)} = (R^{(1)}_x)_{x\in E}$ be a random kernel on E, i.e. $R^{(1)}_x(E)$ is a finite measure on E, for every $x \in E$. We define the Measure-Valued Pólya Process of initial composition m_0 and replacement kernel $R^{(1)}$ recursively as follows: for every $n \geq 0$, 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 $\frac{m_n}{m_n(E)}$ and given $Y(n+1)$, $R^{(n+1)}$ is an independent copy of $R^{(1)}_{Y(n+1)}$.

In the case of Pólya urns, E denotes the space of colours. For all Borel sets $B \subset E$ and integers $n \geq 0$, $m_n(B)$ represents the mass of the balls in the urn at time n, whose colour belongs to B. This means that now the balls are not necessarily of mass 1 anymore, but they can have infinitesimal mass. For every $n \geq 0$, $Y(n)$ denotes the colour of the ball drawn at time n.

Example 4.2 (**The finitely many colour Pólya urn**)**.** Let $E = \{1, \dots, d\}$ for some $d \geq 2$. Consider a matrix $\mathcal{R} = (\mathcal{R}_{i,j})_{1\leq i,j\leq d}$ and a vector $\alpha = (\alpha_1, \dots, \alpha_d)$ both integer-valued. Let $(m_n)_{n\geq 0}$ 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 $R_i = \sum_{j=1}^d \mathcal{R}_{i,j}\delta_j$. Then $m_n \overset{\mathrm{d}}{=} \sum_{i=1}^d U_i(n)\delta_i$ for all $n \geq 0$, where $(U(n))_{n\geq 0}$ is the Pólya urn of initial composition α and replacement matrix $\mathcal{R}$. Hence, all finitely many colour Pólya urns are MVPPs.

4.1 The identity kernel case

In the following theorem, Blackwell and McQueen [BM73] demonstrate that in the identity case, where $R_x = \delta_x$ for all $x \in E$, the composition measure m_n, once renormalized, converges to a limiting random measure. The distribution of this limiting random measure is known as the Ferguson distribution, which we now define:

Definition 4.3 (**Ferguson distribution**)**.** Let μ be a finite measure on E. We say that a random measure on E has *Ferguson distribution of parameter* μ if for every Borel set partition $B_1, \dots, B_d$, $d \geq 1$, $(\mu^*(B_1), \dots, \mu^*(B_d)) \sim \mathrm{Dir}(\mu(B_1), \dots, \mu(B_d))$.

Remark 4.4. A Ferguson-distributed random measure is almost surely discrete, as shown in [Fer73].

Theorem 4.5. *Let μ be a finite measure on a Polish space E. Let $(m_n)_{n\geq 0}$ be the MVPP of initial composition μ and deterministic replacement kernel $Id = (\delta_x)_{x\in E}$. Then,*

$$\hat{m}_n \coloneqq \frac{m_n}{m_n(E)} \xrightarrow{n\to\infty} \mu^* \text{ a.s.},$$

where $\mu^* \sim$ Ferguson(μ).

Proof. We define $(m_n)_{n\geq 0}$ as follows: let $m_0 = \mu$, and for every $n \geq 0$, given m_n, we sample an integer in $\{0, 1 \dots, n\}$ with probabilities

$$\mathbb{P}(I(n+1) = i) = \begin{cases} \frac{\mu(E)}{\mu(E)+n} & \text{for } i = 0 \\ \frac{1}{\mu(E)+n} & \text{for } 1 \leq i \leq n \end{cases}$$

So, if $I(n+1) = 0$, then sample $Y(n+1)$ according to μ, and if $I(n+1) = i$ for some $1 \leq i \leq d$, then set $Y(n+1) = Y(i)$. One can check that, indeed, $(m_n)_{n\geq 0}$ is the MVPP of initial composition μ and replacement kernel Id.
The process of identifying new colours sampled from the distribution μ is clarified by examining the times n at which $I(n) = 0$. These times correspond to moments when a new colour is introduced, contrasting with other times when an already present colour is reinforced. If μ has atoms (discrete points with positive probability), it is possible for a new colour to coincide with an existing one. To manage this, we can extend the set of colours to $E \times \mathbb{N}$. This way, if a colour $x \in E$ is picked more than once, it is distinguished by representing its subsequent occurrences as $(x, 2)$, $(x, 3)$, and so on, while its first occurrence is denoted as $(x, 1)$. However, for simplicity, we assume that μ has no atoms, ensuring each sampled colour is always distinct. We denote the index of the i-th 0 in the sequence $(I(n))_{n\geq 1}$ as ν_i, marking the i-th time a new colour is sampled. Consequently, set $\xi_i = Y(\nu_i)$ for all $i \geq 1$, forming a sequence $(\xi_i)_{i\geq 1}$ of independent and identically distributed (i.i.d.) random variables with distribution μ. This sequence accurately represents the introduction of new colours each time $I(n) = 0$, ensuring the process is well-defined and facilitating subsequent analysis in the proof.
Consider the process $N_1(n) = \#\{1 \leq i \leq n : Y(i) = Y(1)\}$. Note that $(N_1(n), n + \mu(E) - N_1(n))_{n\geq 1}$ is a Pólya urn of initial composition $(1, \mu(E))$ and replacement matrix

Id. Remark that $\mu(E)$ might not necessarily be an integer, but one can check that all the results from chapter 3 still apply. Applying Corollary 3.6(ii), we get that

$$\frac{N_1(n)}{\mu(E)+n} \xrightarrow{n\to\infty} V_1 \text{ a.s.},$$

where $V_1 \sim \text{Beta}(1, \mu(E))$. This implies that

$$\frac{N_1(n)}{n} \xrightarrow{n\to\infty} V_1 \text{ a.s..}$$

Notice that

$$\sum_{n\geq 1} \mathbb{P}(I(n)=0) = \sum_{n\geq 0} \mathbb{P}(I(n+1)=0) = \sum_{n\geq 0} \frac{\mu(E)}{\mu(E)+n} = +\infty,$$

and so by the Borel-Cantelli Lemma (Theorem 2.5), it follows that

$$\mathbb{P}(\limsup_{n\to\infty}\{I(n)=0\}) = 1,$$

i.e. the time τ_2 at which a second colour is introduced is almost surely finite.
Consider now $N_2(n) = \#\{1 \leq i \leq n : Y(i) = \xi_2\}$. Note that $(N_2(n), n+\mu(E) - N_1(n) - N_2(n))_{n\geq 1}$, reduced to its jump-times is also a Pólya urn with initial composition $(1, \mu(E))$ and replacement matrix Id. Thus, it also follows that

$$\frac{N_2(n)}{n+\mu(E)-N_1(n)} \xrightarrow{n\to\infty} V_2 \text{ a.s.},$$

where $V_2 \sim \text{Beta}(1, \mu(E))$ and is independent of V_1. So,

$$\frac{N_2(n)}{n} = \frac{N_2(n)}{n+\mu(E)-N_1(n)} \cdot \frac{n+\mu(E)-N_1(n)}{n} \xrightarrow{n\to\infty} V_2(1-V_1) \text{ a.s..}$$

We can iterate this argument, to get, almost surely, for all $k \geq 1$,

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

Such a sequence $(P_k)_{k\geq 1}$ can be interpreted as a partition of the unit interval; in this context, since $(V_i)_{i\geq 1}$ is a sequence of i.i.d. random variables Beta-distributed of parameter $(1, \mu(E))$, the distribution $(P_k)_{k\geq 1}$ is called *GEM* (after Griffiths, Engen and McCloskey) of parameter $\mu(E)$. Now, for all $n \geq 1$, we can write

$$m_n = \mu + \sum_{k\geq 1} N_k(n)\delta_{\xi_k}, \tag{4.1}$$

where we set $N_k(n) = 0$ if $\#\{Y(1), \dots, Y(n)\} < k$. Hence, for every function $f : E \to \mathbb{R}$, continuous and bounded, we have

$$\begin{aligned}\int_E f \,\mathrm{d}\hat{m}_n &= \frac{1}{m_n(E)} \int_E f \,\mathrm{d}m_n \\ &= \frac{1}{\mu(E)+n}\bigg(\int_E f \,\mathrm{d}\mu + \sum_{k\geq 1} N_k(n) \int_E f \,\mathrm{d}\delta_{\xi_k} \bigg) \\ &= \frac{1}{\mu(E)+n} \int_E f \,\mathrm{d}\mu + \frac{n}{\mu(E)+n} \sum_{k\geq 1} \frac{N_k(n)}{n} f(\xi_k),\end{aligned}$$

where the second equality follows by the expression we wrote in (4.1).

Claim. $\sum_{k\geq 1} P_k = 1$ *almost surely.*

Proof. We define the summations

$$S_n = \sum_{i=1}^{n} P_i = \sum_{i=1}^{n} V_i \prod_{j=1}^{i-1} (1 - V_j)$$

and

$$R_n = \prod_{j=1}^{n} (1 - V_j).$$

Now we can easily show by induction that $S_n + R_n = 1$, indeed for $n = 1$:

$$S_1 + R_1 = V_1 + (1 - V_1) = 1.$$

Suppose that, for every $n \geq 1$, it holds

$$S_n + R_n = \sum_{i=1}^{n} V_i \prod_{j=1}^{i-1} (1 - V_j) + \prod_{j=1}^{n} (1 - V_j) = 1 \tag{IH}$$

The inductive step $n \mapsto n + 1$ yields

$$\begin{aligned}S_{n+1} + R_{n+1} &= \sum_{i=1}^{n+1} V_i \prod_{j=1}^{i-1} (1 - V_j) + \prod_{j=1}^{n+1} (1 - V_j) \\ &= \sum_{i=1}^{n} V_i \prod_{j=1}^{i-1} (1 - V_j) + V_{n+1} \prod_{j=1}^{n} (1 - V_j) + (1 - V_{n+1}) \prod_{j=1}^{n} (1 - V_j) \\ &= \sum_{i=1}^{n} V_i \prod_{j=1}^{i-1} (1 - V_j) + (V_{n+1} + 1 - V_{n+1}) \prod_{j=1}^{n} (1 - V_j) \overset{\text{IH}}{=} 1.\end{aligned}$$

Then, since $V_i \sim \mathrm{Beta}(1, \mu(E))$, each $(1 - V_i)$ is a random variable taking values in $(0, 1)$, and thus

$$\mathbb{E}[\log(1 - V_i)] = \int_0^1 \underbrace{\frac{\Gamma(1 + \mu(E))}{\Gamma(1)\Gamma(\mu(E))}}_{=\mu(E)} \cdot \underbrace{\log(1 - x)}_{<0 \text{ for } x\in(0,1)} \cdot (1 - x)^{\mu(E)-1} \,\mathrm{d}x\,.$$

Since $\mu(E)(1-x)^{\mu(E)-1}$ is a valid probability density function for V_i, it follows that $\mathbb{E}[\log(1-V_i)] < 0$. By the law of large numbers (Theorem 2.6), we have

$$\frac{1}{n}\sum_{i=1}^{n}\log(1-V_i) \xrightarrow{n\to\infty} \mathbb{E}[\log(1-V_i)] \text{ a.s.}$$

Hence, $\sum_{i=1}^{n}\log(1-V_i) \sim nc$, for some $c<0$, almost surely. Therefore,

$$R_n = \prod_{j=1}^{n}(1-V_i) = \exp\left\{\log\left(\sum_{j=1}^{n}(1-V_j)\right)\right\} \sim e^{nc} \xrightarrow{n\to\infty} 0 \text{ a.s.}$$

So, $S_n = 1 - R_n \xrightarrow{n\to\infty} 1$ almost surely, i.e. $\sum_{k\geq 1} P_k = 1$ a.s. ■

By the claim, it is possible to fix an $\varepsilon > 0$ and K large enough so that $\sum_{k=1}^{K} P_k \geq 1-\varepsilon$. Then fix n_ε large enough so that $\forall n \geq n_\varepsilon$,

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

The latter assumption is needed also so that $\frac{1}{n+\mu(E)}\int_E f\,\mathrm{d}\mu < \varepsilon$, $\forall n \geq n_\varepsilon$. Then, for every $n \geq n_\varepsilon$,

$$\begin{aligned}
\left|\int_E f\,\mathrm{d}\hat{m}_n - \sum_{k\geq 1} P_k f(\xi_k)\right| &= \left|\frac{1}{n+\mu(E)}\int_E f\,\mathrm{d}\mu + \frac{n}{n+\mu(E)}\sum_{k\geq 1}\left(\frac{N_k(n)}{n} - P_k\right)f(\xi_k)\right| \\
&\leq \varepsilon + \frac{n}{n+\mu(E)}\left|\sum_{k=1}^{K}\left(\frac{N_k(n)}{n} - P_k\right)f(\xi_k) + \sum_{k\geq K}\left(\frac{N_k(n)}{n} - P_k\right)f(\xi_k)\right| \\
&\leq \varepsilon + \underbrace{\sum_{k=1}^{K}\left|\frac{N_k(n)}{n} - P_k\right|}_{\leq\varepsilon}\|f\|_\infty + \sum_{k\geq K}\left|\frac{N_k(n)}{n} - P_k\right|\|f\|_\infty \\
&\leq (1+\|f\|_\infty)\varepsilon + \|f\|_\infty \sum_{k\geq K}\frac{N_k(n)}{n} + \|f\|_\infty \underbrace{\sum_{k\geq K} P_k}_{\leq\varepsilon} \\
&\leq (1+2\|f\|_\infty)\varepsilon + \|f\|_\infty\left(1 - \sum_{k=1}^{K}\frac{N_k(n)}{n}\right) \\
&\leq (1+4\|f\|_\infty)\varepsilon,
\end{aligned}$$

where the last inequality follows from the fact that

$$\sum_{k=1}^{K} \frac{N_k(n)}{n} = \sum_{k=1}^{K} \left(\frac{N_k(n)}{n} - P_k \right) + \underbrace{\sum_{k=1}^{K} P_k}_{\geq 1-\varepsilon} \geq -\sum_{k=1}^{K} \left| \frac{N_k(n)}{n} - P_k \right| + 1 - \varepsilon$$

$$\geq -\underbrace{\left| \sum_{k=1}^{K} \frac{N_k(n)}{n} - P_k \right|}_{\leq \varepsilon} + 1 - \varepsilon \geq 1 - 2\varepsilon.$$

Now, because ε can be arbitrarily small, this shows that

$$\int_E f \, \mathrm{d}\hat{m}_n \xrightarrow{n\to\infty} \sum_{k\geq 1} P_k f(\xi_k) \text{ a.s.} \tag{4.2}$$

We still need to show that $\mu^* \coloneqq \sum_{k\geq 1} P_k f(\xi_k) \sim \text{Ferguson}(\mu)$. This follows from Theorem 3.4, i.e. the finitely many colour case: indeed, for any partition $(B_1, \ldots, B_d)$ of E, for every $i \geq 1$ let

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

in other words $X(i) = e_k$ if $Y(i) \in B_k$. Then, for every $n \geq 0$, we define

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

and thus

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

Notice that, for all $n \geq 0$,

$$\mathbb{P}(X(n+1) = k \big| U(n)) = \mathbb{P}(Y(n+1) \in B_k \big| U(n)) = \mathbb{E}[\mathbb{P}(Y(n+1) \in B_k \big| U(n), m_n) \big| U(n)],$$

where the last step follows from the law of total expectation (see also equation (3.4)), and we condition on both the current state $U(n)$ and the measure m_n, which together give a complete picture of the system's state. By conditioning also on m_n, we leverage the accumulated knowledge about which sets B_k the balls have been drawn from. This allows us to calculate the probability of the next ball being of a certain colour, given the detailed history encapsulated in m_n. So,

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

In words, this means that $U(n)$ is the Pólya urn of initial composition $(\mu(B_1), \dots, \mu(B_d))$ and replacement matrix Id. Applying Theorem 3.4 to the almost sure limit $\frac{U(n)}{n}$, which is μ^* as we found in equation 4.2, we get that $(\mu^*(B_1), \dots, \mu^*(B_d))$ is Dirichlet distributed of parameter $(\mu(B_1), \dots, \mu(B_d))$, i.e. by Definition 4.3, μ^* is Ferguson distributed of parameter μ. □

Blackwell and McQueen [BM73] also prove that Theorem 3.7 still holds in this case:

Theorem 4.6. *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 $(X(i))_{i\geq 1}$ be a sequence of i.i.d. random variables of distribution μ^*, and let, for all $n \geq 0$,*

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

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

Proof. We follow the argument of [FPS21]. First we define what is an exchangeable sequence of random variables (cf. Example 4.7.8 of [Dur19]): a sequence $(X_n)_{n\geq 1}$ is said to be exchangeable if for each n and permutation π of $\{1, \dots, n\}$, $(X_1, \dots, X_n)$ and $(X_{\pi(1)}, \dots, X_{\pi(n)})$ have the same distribution. *De Finetti's theorem* states that, if $(X_n)_{n\geq 1}$ is a sequence of exchangeable random variables, then there exists a random probability measure $\tilde{\mathbb{P}}$ on E such that, given $\tilde{\mathbb{P}}$, the random variables $X_1, X_2, \dots$ are conditionally independent and identically distributed with marginal distribution $\tilde{\mathbb{P}}$ (see section 3 of [AIJA85], Theorem 4.7.9 and 4.7.10 of [Dur19]). Furthermore, $\tilde{\mathbb{P}}$ is the almost sure limit of the predictive distributions and the empirical frequencies, that is, given $(X_i)_{1\leq i\leq n}$, $X_{n+1} \sim \tilde{\mathbb{P}}$, and

$$\frac{1}{n}\sum_{i=1}^{n} \delta_{X_i} \xrightarrow{n\to\infty} \tilde{\mathbb{P}}.$$

By assumption on the sequence $(X(i))_{i\geq 1}$, it is straightforward to see that an i.i.d. sequence is also an exchangeable sequence, thus $(X(i))_{i\geq 1}$ is a sequence of exchangeable random variables. So, by the arguments above, we can identify this probability measure $\tilde{\mathbb{P}}$ with our $\mu^* \sim$ Ferguson(μ). Hence, by de Finetti's theorem and the arguments in the proof of Theorem 4.5, this prove the theorem.

□

For the variations of $\hat{m}_n$ around its almost certain limit μ^*, Theorem 4.6 can be applied. See [Bor21] for a functional limit theorem regarding these fluctuations.

4.2 The balanced irreducible case

In this section we assume that the MVPP is *balanced*, which means that there exists $c > 0$ such that, for every $x \in E$, $R_x(E) = c$. In the case of finitely many colours Pólya urns, this would imply that, for a Pólya urn $(U(n))_{n\geq 0}$, the total number of balls in

the urn at time n is $nc + U(0)$, but this assumption was not necessary for the results we derived in chapter 3, although other studies of standard Pólya urn processes assume this condition. Without loss of generality, we assume that $c = 1$, which corresponds to saying that, for every $x \in E$, R_x is a probability measure: we call R a *probability kernel*. The non-balanced case will be treated in the next section.

4.2.1 Convergence in probability of irreducible MVPPs

Definition 4.7. Let $(a(n))_{n\geq 0}$, $(b(n))_{n\geq 0}$ be two E-valued sequences and let ν be a probability distribution on E. We say that the probability kernel $(R_x)_{x\in E}$ on E is (a, b, ν)-*ergodic* if, for all $x \in E$, the Markov chain $(W(n))_{n\geq 0}$ started at x and whose transition probabilities are given by R, that is $W(n+1) \sim R_{W(n)}$, for every $n \geq 0$, satisfies

$$\frac{W(n) - a(n)}{b(n)} \xrightarrow{d} \nu \quad \text{as } n \to \infty.$$

Remark 4.8. (i) If E is not equipped with an addition, then $(a(n))_{n\geq 0}$ can only be the constant sequence equal to 0, interpreting $x \mapsto x + 0$ as the identity function on E.

(ii) If E is not equipped with a multiplication by a scalar, then $(b(n))_{n\geq 0}$ is the constant sequence equal to 1, and we interpret $x \mapsto 1 \times x$ as the identity function on E.

Theorem 4.9. *Let $(m_n)_{n\geq 0}$ be the MVPP of initial composition m_0 and balanced replacement kernel $(R_x)_{x\in E}$. Assume there exists $a = (a(n))_{n\geq 0}$, $b = (b(n))_{n\geq 0}$ 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 every $x \in E$, for any function $(\varepsilon_n)_{n\geq 0}$ satisfying $\varepsilon_n = o(\sqrt{n})$ as $n \to \infty$,*

$$\frac{a(n + x\sqrt{n} + \varepsilon_n) - a(n)}{b(n)} \to f(x) \quad \textit{and} \quad \frac{b(n + x\sqrt{n} + \varepsilon_n)}{b(n)} \to g(x), \tag{4.3}$$

as $n \to \infty$. Then, as $n \to \infty$,

$$\hat{m}_n(a(\log(n)) + \cdot\, b(\log(n))) \xrightarrow{\mathbb{P}} \mathcal{L}(f(\Lambda) + g(\Lambda)\Gamma),$$

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

Remark 4.10. The latter convergence means that, for any continuous bounded function $\phi : E \to \mathbb{R}$, as $n \to \infty$,

$$\int_E \phi\left(\frac{x - a(\log(n))}{b(\log(n))}\right) \mathrm{d}\hat{m}_n(x) \xrightarrow{\mathbb{P}} \mathbb{E}\big[\phi(f(\Lambda) + g(\Lambda)\Gamma)\big].$$

Before we prove the latter theorem, we have a look at a couple of examples of applications of the theorem.

Example 4.11 (Finitely-many colour Pólya urn)**.** We recall the notions already given in Example 4.2: let $(U(n))_{n\geq 0}$ be the d-colour Pólya urn, for some $d \geq 2$, with replacement matrix $\mathcal{R} = (\mathcal{R}_{i,j})_{1\leq i,j\leq d}$ and initial composition $\alpha = (\alpha_1, \ldots, \alpha_d)$. We argued that $(m_n = \sum_{i=1}^{d} U(n)\delta_i)_{n\geq 0}$ is the MVPP of initial composition $m_0 = \sum_{i=1}^{d} \alpha_i \delta_i$ and replacement kernel $(R_i = \sum_{j=1}^{d} \mathcal{R}_{i,j}\delta_j)_{1\leq i\leq d}$. In order to apply Theorem 4.9, we need the, so-called in the literature, balance assumption, which means that, for every $1 \leq i \leq d$, $\sum_{j=1}^{d} \mathcal{R}_{i,j} = \mathcal{S}$, for some $\mathcal{S} \geq 1$. For all $n \geq 0$, let $\eta_n = \frac{m_n}{\mathcal{S}}$, then $(\eta_n)_{n\geq 0}$ is a balanced MVPP of replacement kernel $\frac{R}{\mathcal{S}}$ and initial composition $\frac{\alpha}{\mathcal{S}}$. If we also assume that $\mathcal{R}$ is irreducible and aperiodic, then the Markov chain is ergodic (positive recurrence is also implied, since the Polish space E is finite in this case). Then, its limiting probability distribution satisfies $v \cdot \frac{R}{\mathcal{S}} = v$, which we can write as $vR = \mathcal{S}v$. So, we can apply Theorem 4.9 and we get

$$\frac{U(n)}{\|U(0)\|_1 + n\mathcal{S}} \xrightarrow{\mathbb{P}} v,$$

as $n \to \infty$. Since we assumed that R is balanced, all its rows sum up to $\mathcal{S}$, hence $\mathcal{S}$ is its dominant eigenvalue, which implies that v is its dominant left-eigenvector. Thus, we recover a version of Theorem 3.14, but this version is weaker, since we had to assume balance and aperiodicity of the Pólya urn, and moreover we get convergence in probability instead of almost surely.

Example 4.12 (Random walk increment)**.** Let $(m_n)_{n\geq 0}$ 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 being a random variable on $\mathbb{R}^d$ such that $\mathbb{E}[\Delta] = \mu$ and $\mathrm{Var}(\Delta) = \sigma^2 < \infty$. Then, the Markov chain $(W(n))_{n\geq 0}$ of probability transitions $(R_x)_{x\in\mathbb{R}^d}$ is the simple random walk of increment Δ. This means that $W(n) = W(0) + \sum_{i=1}^{n} \Delta_i$, for all $n \geq 0$, where $(\Delta_i)_{i\geq 1}$ is a sequence of i.i.d. copies of Δ. Applying the central limit theorem (Theorem 2.8), we get that $W(n)$ is (a, b, ν)-ergodic, for $a(n) = \mu n$, $b(n) = \sqrt{n}$ and $\nu = \mathcal{N}(0, \sigma^2)$. Note that this satisfies conditions 4.3, with $f(x) = \mu x$ and $g(x) = 1$. So, by Theorem 4.9,

$$\frac{m_n}{m_n(E)} \xrightarrow{\mathbb{P}} \mathcal{L}(\mu\Omega + \Gamma),$$

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

$$\frac{m_n}{m_n(E)} \xrightarrow{\mathbb{P}} \mathcal{N}(0, \mu^2 + \sigma^2).$$

4.2.2 Random recursive trees and branching Markov chains

We now introduce some new notions, which will facilitate us in proving Theorem 4.9, starting by giving the following, equivalent, definition of an MVPP. Given a replacement kernel R and an initial composition m_0, the sequence $(m_n)_{n\geq 0}$ is defined as follows: first, sample a sequence $(I(n))_{n\geq 1}$ of independent random variables such that, for every $n \geq 1$,

$$\mathbb{P}(I(n) = i) = \begin{cases} \frac{m_0(E)}{m_0(E)+n-1} & \text{for } i = 0 \\ \frac{1}{m_0(E)+n-1} & \text{for } 1 \leq i \leq n-1 \end{cases}$$

Recalling that $Y(i)$ denotes the colour of the ball drawn at time i, for some $1 \leq i \leq n$, for every $n \geq 0$, given $Y(1), \ldots, Y(n)$, we sample $Y(n+1)$ as follows:

- if $I(n+1) = i$ for some $1 \leq i \leq n$, then we sample $Y(n+1)$ according to the probability distribution $R_{Y(i)}$;
- if $I(n+1) = 0$, we sample $Y(n+1)$ according to m_0.

Further, we set $m_n = m_0 + \sum_{i=1}^{n} R_{Y(i)}$, for all $n \geq 0$. So, $(m_n)_{n\geq 0}$ is the MVPP of replacement kernel R and initial composition m_0.
This definition is useful, as the sequence $(I(n))_{n\geq 1}$ records the branching structure of the MVPP. Specifically, $I(n)$ can be interpreted as the parent of n, for all $n \geq 1$. With this definition, all positive integers have a parent, and the genealogical structure of this population can be interpreted as a tree. Each integer is interpreted as a node, and we add an edge between every node and its parent (since the parent is a smaller integer, the graph contains no cycles). The only node without a parent is 0, which is addressed as the root of the tree. The tree T_∞ associated to $(I(n))_{n\geq 1}$ is infinite; for every $n \geq 1$, T_n denotes the tree associated to the finite sequence $(I(1), \ldots, I(n))$. Since $(I(n))_{n\geq 1}$ is a random sequence, $(T_n)_{n\geq 0}$ is a sequence of random trees.
Assuming $m_0(E) = 1$ (for simplicity), then $(T_n)_{n\geq 0}$ is known as the *random recursive tree* (RRT). Indeed, we can confront this with the definition given in [Mah08] (section 8.2):

Definition 4.13 (RRT)**.** Starting with a single root node, the random recursive tree $(T_n)_{n\geq 0}$ grows in discrete time steps. At each step, a node in the tree is chosen uniformly at random to be the parent of a new entrant. If the n-th node is labelled with n, all root-to-leaf paths will correspond to increasing sequences. There is no restriction on node degrees in recursive trees. and they can grow indefinitely.

As an example, for $n = 3$, we have the following figure:

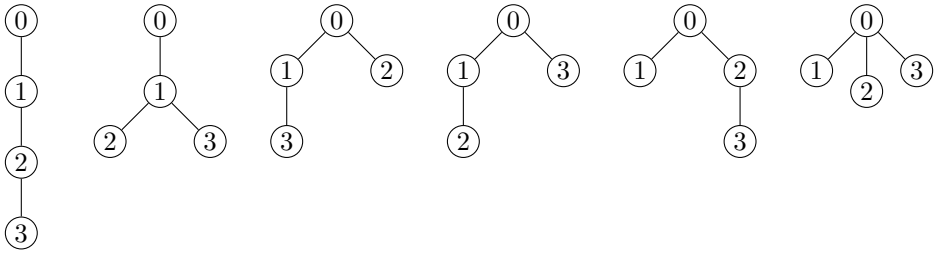

Figure 4.1: All possible recursive trees T_n for $n = 3$.

The sequence $(Y(i))_{i\geq 0}$ of colours drawn in the MVPP can be interpreted, not as a sequence indexed by $\mathbb{N}$, but as a sequence indexed by the RRT T_∞. We now 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 with initial distribution m_0 and transition probabilities R. And once they have branched, the labels along two distinct branches are two independent Markov chains with transition probabilities R. Thus, $(Y(i))_{i\geq 0}$, seen as a sequence indexed by T_∞, is a branching Markov chain.

The benefit of this description is the following result:

Lemma 4.14. *Let R be a replacement kernel on a Polish space E and m_0 a finite measure on E. Assume that $R_x(E) = 1$ for every $x \in E$. Let T_∞ be the random recursive tree, whose nodes are $\{0, 1, 2, \dots\}$, and for all $n \geq 0$, let T_n be the restriction of T_∞ to $\{0, \dots, n\}$. Let $(Y(i))_{i\geq 1}$ be the branching Markov chain of initial distribution m_0 and transition probabilities R indexed by T_∞. Then, for every $n \geq 0$, given T_n and $(Y(i))_{1\leq i\leq n}$, $\hat{m}_n = \frac{m_n}{m_n(E)}$ is the distribution of $Y(n+1)$.*

Proof. This is Lemma 18 of [MM17] and we follow its proof.
To establish the lemma, it is enough to show that

$$m_n^* = m_0 + \sum_{i\in T_n\setminus\{0\}} R_{Y(i)}$$

satisfies $(m_n^*)_{n\geq 0} = (m_n)_{n\geq 0}$, where $(m_n)_{n\geq 0}$ is the MVPP of initial composition m_0 and replacement kernel $(R_x)_{x\in E}$ (see Definition 4.1). As previously mentioned, constructing T_{n+1} from T_n involves selecting a node $j \in T_n$ uniformly at random and adding a new child $n+1$ to j. Consequently, in the branching Markov chain, the distribution of the new label $Y(n+1)$ depends only on the already existing labels $(Y(i), i \in T_n)$, and not on the tree's geometry (indeed, one can see in the example depicted in Figure 4.1 all possible recursive trees for $n = 3$, as they do not depend on the geometry of the tree). This ensures that m_n^* is a Markov chain. The initial composition of $(m_n^*)_{n\geq 0}$ and the one of $(m_n)_{n\geq 0}$ are the same, that is m_0, by definition of the processes. To show that they have the same replacement kernel, note that the only distinction between the MVPP and the branching Markov chain representation $m_0 + \sum_{i\in T_n\setminus\{0\}} R_{Y(i)}$ lies in the arrangement of the current values Y. However, the measures m_n^* do not depend on this arrangement; this concludes the proof that $(m_n^*)_{n\geq 0} = (m_n)_{n\geq 0}$. □

Thanks to this coupling, we proceed to prove Theorem 4.9.

4.2.3 Proof of Theorem 4.9

We now just focus on the branching Markov chain on the random recursive tree, introduced in the previous section. As we discussed, in particular in Lemma 4.14, we have that $Y(n+1) \overset{(d)}{=} W_{|n+1|}$, where W is the Markov chain of initial distribution m_0 and replacement probabilities R and $|n+1|$ denotes the height of node $n+1$ in T_∞. Our assumptions provide good control over $W_{|n+1|}$, given that $|n+1|$ approaches infinity. The upcoming result allows us control over $|n+1|$ as n tends to infinity; instead of focusing on node $n+1$, we consider its (random) parent $\xi(n+1)$. Note that $\xi(n+1)$ is uniformly distributed over $\{0, \dots, n\}$.

Lemma 4.15 (Dobrow, [Dob96])**.** *Let T_n be the random recursive tree. Let $\xi(n+1)$ be a node taken uniformly at random among $\{0, 1 \dots, n\}$. Then, the height of $\xi(n+1)$, which $|\xi(n+1)|$ denotes, satisfies, as $n \to \infty$,*

$$\frac{|\xi(n+1)| - \log(n)}{\sqrt{\log(n)}} \xrightarrow{d} \mathcal{N}(0, 1). \tag{4.4}$$

Proof. This results was initially established by Mahmoud ([Mah91]) and later generalized by Dobrow ([Dob96]). Mahmoud's approach involves the use of moment generating function and utilizing the known distribution of the random variable $|\xi(n+1)|$, as determined by Szymański ([Szy90]). Szimański's work connects this distribution to Stirling numbers of the first kind. For brevity, the details are omitted. □

Recall that, by assumption, $(R_x)_{x\in E}$ is (a, b, ν)-ergodic, i.e.

$$\frac{W_n - a_n}{b_n} := \frac{W(n) - a(n)}{b(n)} \xrightarrow{d} \nu, \tag{4.5}$$

as $n \to \infty$. By Skorohod's representation theorem (Theorem 2.10), there exists a probability space, referred to as *Skorohod's space* hereafter, where this convergence (4.5) and the convergence of (4.4) occur simultaneously almost surely (for simplicity, we use the same notation, although we are actually dealing with different objects on a different probability space):

$$\left(\frac{|\xi(n+1)| - \log(n)}{\sqrt{\log(n)}}, \frac{W_n - a_n}{b_n}\right) \xrightarrow{n\to\infty} (\Omega, \Gamma) \quad \text{a.s.},$$

where $\Omega \sim \mathcal{N}(0, 1)$ and $\Gamma \sim \nu$ are independent.
Now, notice that

$$\begin{aligned}\frac{W_{|n+1|} - a_{\log(n)}}{b_{\log(n)}} &= \frac{W_{|\xi(n+1)|+1} - a_{|\xi(n+1)|+1}}{b_{|\xi(n+1)|+1}} \cdot \frac{b_{|\xi(n+1)|+1}}{b_{\log(n)}} + \frac{a_{|\xi(n+1)|+1} - a_{\log(n)}}{b_{\log(n)}} \\ &\xrightarrow{n\to\infty} \Gamma g(\Omega) + f(\Omega) \quad \text{a.s.}\end{aligned} \tag{4.6}$$

However, the almost sure convergence is true only on Skorohod's space, but this implies convergence in distribution on the original probability space. We thus proved that,

$$\frac{W_{|n+1|} - a_{\log(n)}}{b_{\log(n)}} \xrightarrow{d} \mathcal{L}(\Gamma g(\Omega) + f(\Omega)),$$

as $n \to \infty$. Equivalently, for any bounded continuous function $\phi : E \to \mathbb{R}$,

$$\mathbb{E}\left[\phi\left(\frac{W_{|n+1|} - a_{\log(n)}}{b_{\log(n)}}\right)\right] \xrightarrow{n\to\infty} \mathbb{E}\big[\phi(\Gamma g(\Omega) + f(\Omega))\big].$$

But, even if $(W_{|n+1|} - a_{\log(n)})/b_{\log(n)}$ is a random variable with distribution equal to $\hat{m}_n(a_{\log(n)} + \cdot b_{\log(n)})$, this alone does not suffice to prove convergence in probability of $\hat{m}_n(a_{\log(n)} + \cdot b_{\log(n)})$ to $\mathcal{L}(\Gamma g(\Omega) + f(\Omega))$. To establish this, we require additional conditions, as stated in the following result:

Lemma 4.16. *Let $(\mu_n)_{n\geq 0}$ be a sequence of random probability distributions on a Polish space E. Let μ be a deterministic probability distribution on E. For every $n \geq 0$, given μ_n, let A_n and B_n be two independent random variables of distribution μ_n. If $(A_n, B_n) \xrightarrow{d} (A, B)$ as $n \to \infty$, where A and B are two independent random variables*

of distribution μ, then $\mu_n \xrightarrow{\mathbb{P}} \mu$ for the weak topology. I.e., for any bounded continuous function $\phi : E \to \mathbb{R}$, as $n \to \infty$,

$$\int_E \phi \, \mathrm{d}\mu_n \xrightarrow{\mathbb{P}} \int_E \phi \, \mathrm{d}\mu \,.$$

Proof. Since μ is deterministic, it suffices to prove that, for any bounded continuous function $\phi : E \to \mathbb{R}$,

$$\mathbb{E}\bigg[\int_E \phi \, \mathrm{d}\mu_n \bigg] \xrightarrow{n\to\infty} \int_E \phi \, \mathrm{d}\mu \quad \text{and} \quad \mathrm{Var}\bigg(\int_E \phi \, \mathrm{d}\mu_n \bigg) \xrightarrow{n\to\infty} 0.$$

First, notice that, by definition of A_n, it holds $\int_E \phi \, \mathrm{d}\mu_n = \mathbb{E}[\phi(A_n)|\mu_n]$, thus

$$\mathbb{E}\bigg[\int_E \phi \, \mathrm{d}\mu_n \bigg] = \mathbb{E}[\mathbb{E}[\phi(A_n)|\mu_n]] = \mathbb{E}[\phi(A_n)] \xrightarrow{n\to\infty} \mathbb{E}[\phi(A)] = \int_E \phi \, \mathrm{d}\mu \,,$$

where the convergence follows by the assumption that $A_n \xrightarrow{d} A$. Then, also since $\mathbb{E}[B_n|\mu_n] = \mathbb{E}[A_n|\mu_n] = \int_E \phi \, \mathrm{d}\mu_n$, and given μ_n, A_n and B_n are independent, we have

$$\begin{aligned} \mathbb{E}\bigg[\bigg(\int_E \phi \, \mathrm{d}\mu_n \bigg)^2 \bigg] &= \mathbb{E}[\mathbb{E}[A_n|\mu_n]\mathbb{E}[B_n|\mu_n]] = \mathbb{E}[\mathbb{E}[A_nB_n|\mu_n]] = \mathbb{E}[A_nB_n] \\ &\xrightarrow{n\to\infty} \mathbb{E}[AB] = \mathbb{E}[A]\mathbb{E}[B] = \bigg(\int_E \phi \, \mathrm{d}\mu \bigg)^2, \end{aligned}$$

since A and B are independent. This shows that $\mathrm{Var}(\int_E \phi \, \mathrm{d}\mu_n) \xrightarrow{n\to\infty} 0$. So, by Chebyshev's inequality (Corollary 2.4), this implies that

$$\int_E \phi \, \mathrm{d}\mu_n \xrightarrow{\mathbb{P}} \int_E \phi \, \mathrm{d}\mu \,.$$

□

Thus, to prove Theorem 4.9, it is enough to prove the following result:

Proposition 4.17. *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 RRT and let $(Y(i))_{1\le i\le n}$ be the branching Markov chain of initial distribution m_0 and transition probabilities R indexed by T_∞. Given T_n and $(Y(i))_{1\le i\le d}$, let $Y_1(n+1)$ and $Y_2(n+1)$ be two independent copies of $Y(n+1)$. Then, as $n \to \infty$,*

$$\bigg(\frac{Y_1(n+1) - a_{\log(n)}}{b_{\log(n)}}, \frac{Y_2(n+1) - a_{\log(n)}}{b_{\log(n)}} \bigg) \xrightarrow{d} \big(f(\Omega_1) + g(\Omega_1)\Gamma_1, f(\Omega_2) + g(\Omega_2)\Gamma_2 \big),$$

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 improve Lemma 4.15 as follows:

Lemma 4.18. *Let T_∞ be the random recursive tree. Let $\zeta(n+1)$ and $\xi(n+1)$ be two nodes taken independently uniformly at random in $\{0,1,\dots,n\}$. Let $|\zeta(n+1)|, |\xi(n+1)|$ denote their respective heights in T_∞, and let K_n be the height of their last common ancestor in T_∞. Then, as $n \to \infty$,*

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

where $\Omega_1, \Omega_2 \sim \mathcal{N}(0,1)$ are independent, and K is an almost surely finite random variable.

Proof. The proof relies on a coupling that links all the triples $(T_n, \xi(n+1), \zeta(n+1))$ for $n \geq 0$ in such a way that, for every $n \geq 1$,

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

We proceed in this manner: for T_0, let $\xi(1) = \zeta(1) = 0$, i.e. the root of the tree. Then, we sample two independent sequences of independent random variables $(B_i)_{i\geq 1}$ and $(B'_i)_{i\geq 1}$ Bernoulli distributed of parameter $1/i$, that is

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

We also sample a sequence $(U_n)_{n\geq 1}$ of independent random variables such that, for all $n \geq 1$, U_n is uniform on $\{0,\dots,n-1\}$. Then, given $(T_{n-1}, \xi(n), \zeta(n))$, we define $(T_n, \xi(n+1), \zeta(n+1))$ as follows:

- if $B_n = B'_n = 1$, then set $\zeta(n+1) = \xi(n+1) = n$, and T_n is the tree resulting 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 $\zeta(n+1) = n$, $\xi(n+1) = \xi(n)$, and T_n is the tree resulting 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 $\zeta(n+1) = \zeta(n)$, $\xi(n+1) = n$, and T_n is the tree resulting after adding node n to T_{n-1} as a child of node $\xi(n)$;
- if $B_n = B'_n = 0$, then set $\zeta(n+1) = \zeta(n)$, $\xi(n+1) = \xi(n)$, and T_n is the tree resulting after adding node n to T_{n-1} as a child of node U_n.

It can be verified that $(T_n)_{n\geq 0}$ is indeed distributed as the random recursive tree, and for every $n \geq 0$, $\zeta(n+1)$ and $\xi(n+1)$ are two nodes chosen uniformly at random from $\{0,\dots,n\}$. In [MB19], a proof of this is provided for the weighted random recursive tree, a generalization of our RRT.
Notice that, when n evolves, $\zeta(n+1)$ stays on a single branch of the tree and occasionally increases its height $|\zeta(n+1)|$ by one (when $B_n = 1$). In contrast, $\xi(n+1)$ moves along one branch of the tree, sometimes increasing its height by one (if $B'_n = 1$ and $B_n = 0$). However, at certain random times (when $B_n = B'_n = 1$), $\xi(n+1)$ "jumps" to the same

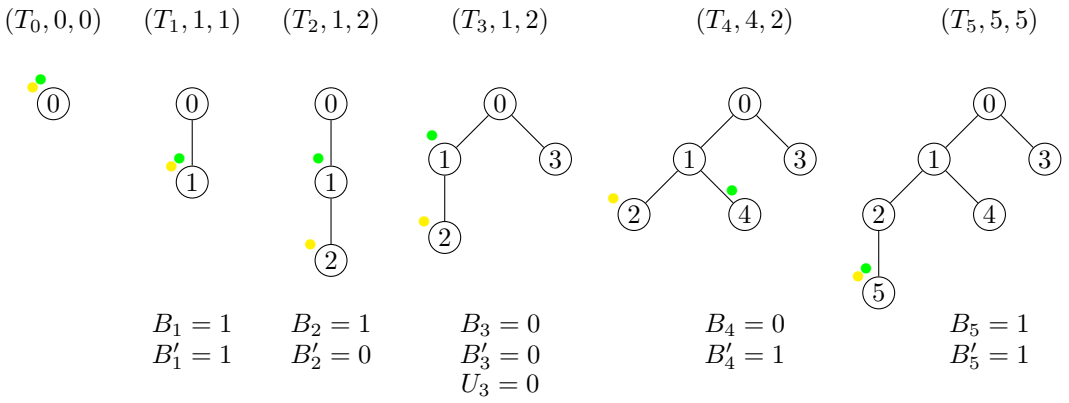

Figure 4.2: A realization of the triples $(T_n, \xi(n+1), \zeta(n+1))$ for $0 \leq n \leq 5$. For each n, the green dot marks the node equal to $\xi(n+1)$ in T_n, the yellow dot marks the node equal to $\zeta(n+1)$.

position as $\zeta(n+1)$ (also see the example given in Figure 4.2, where such a jump occurs at $n = 5$). From this behaviour we derive the following identities: for every $n \geq 0$,

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

where $I_n = \max\{1 \leq i \leq n : B_i = B_i' = 1\}$. By Borel-Cantelli lemma (Theorem 2.5), since

$$\sum_{i \geq 1} \mathbb{P}(B_i = B_i' = 1) = \sum_{i \geq 1} \frac{1}{i^2} < \infty,$$

it follows that

$$\lim_{n \to \infty} I_n = \sum_{i \geq 1} B_i B_i' < \infty \quad \text{a.s.}$$

This implies that, $K = \lim_{n \to \infty} K_n < \infty$ almost surely. Then, since

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

and

$$\sum_{i=1}^{n} \mathrm{Var}(B_i) = \sum_{i=1}^{n} \frac{1}{i}\left(1 - \frac{1}{i}\right) = \log(n) + O(1),$$

by the central limit theorem (Theorem 2.8), we get that

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

as $n \to \infty$, where $\Omega_1, \Omega_2 \sim \mathcal{N}(0,1)$ are independent. Now,

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

as $n \to \infty$, by Slutsky's theorem (Theorem 2.7). This concludes the proof. □

Thanks to this, we can now prove Proposition 4.17, and hence Theorem 4.9.

Proof of Proposition 4.17. The idea behind the proof is as follows: $Y_1(n+1)$ represents the label of the first child in T_∞ of a node $\zeta(n+1)$, which is selected uniformly at random among $\{0, \dots, n\}$. Similarly, $Y_2(n+1)$ denotes the label of the second child (if $\xi(n+1) = \zeta(n+1)$) in T_∞ of a node $\xi(n+1)$, chosen uniformly at random among $\{0, \dots, n\}$, independently of $\zeta(n+1)$. We define K_n as the height of the last common ancestor of $\zeta(n+1)$ and $\xi(n+1)$. Then,

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

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, it follows that, as $n \to \infty$,

$$\left(\frac{W^{(1)}_n - a_n}{b_n}, \frac{W^{(2)}_n - a_n}{b_n}\right) \xrightarrow{d} (\Gamma_1, \Gamma_2), \tag{4.7}$$

conditionally on K_n and W_{K_n}, where $\Gamma_1, \Gamma_2 \sim \nu$ are independent. But since the distribution of the limit is independent of K_n and W_{K_n}, the convergence (4.7) holds without conditioning. Furthermore, by Lemma 4.18 we have, as $n \to \infty$,

$$\left(\frac{|\zeta(n+1)| + 1 - K_n - \log(n)}{\sqrt{\log(n)}}, \frac{|\xi(n+1)| + 1 - K_n - \log(n)}{\sqrt{\log(n)}}\right) \xrightarrow{d} (\Omega_1, \Omega_2), \tag{4.8}$$

where $\Omega_1, \Omega_2 \sim \mathcal{N}(0,1)$ are independent. Since the random recursive tree is independent of $W^{(1)}$ and $W^{(2)}$, the convergences (4.7) and (4.8) hold jointly with (Γ_1, Γ_2) being independent of (Ω_1, Ω_2). By Skorohod's representation theorem (Theorem 2.10), there exists a probability space (referred to as Skorohod's space hereafter) on which both convergences (4.7) and (4.8) hold almost surely. The same computation as we did before in (4.6), gives

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

almost surely, on Skorohod's probability space, and therefore in distribution on the original probability space. This ends the proof of the proposition, and thus, also Theorem 4.9 is proved. □

4.3 The non-balanced irreducible case

This section is based on Mailler and Villemonais [MV20] and is dedicated to the unbalanced case. Their paper allows to prove almost sure convergence of unbalanced MVPPs, instead of convergence in probability as in section 4.2, but this does not mean that this result is stronger (nor weaker), as for the unbalanced case only allows the Markov chain of transition probabilities R to be $(0, 1, \nu)$-ergodic. Another difference is the proof method: in the previous section we used what could be called a "many-to-two" method from the theory of branching random walks, while for the unbalanced MVPPs, we use stochastic approximations methods. To eliminate the balance assumption, we rely on the theory of quasi-stationarity for *killed Markov chains*: these refer to Markov processes that are terminated or "killed" at certain times, often modelled by introducing an absorbing state that represents the termination. This allows us to handle the unbalanced nature of the process more effectively.
Before showing the convergence, we introduce stochastic approximations and show that Pólya urns are of this kind.

Definition 4.19 (Stochastic approximation)**.** A stochastic approximation on a subset $S \subset \mathbb{R}^d$ is a sequence $(X_n)_{n\geq 0}$ of S-valued random variables such that, for some filtration $(\mathcal{F}_n)_{n\geq 0}$, for all $n \geq 0$, almost surely,

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

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

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

$F : S \to S$, $(\Delta M_n)_{n\geq 0}$ is an $(\mathcal{F}_n)_{n\geq 0}$-adapted sequence of martingale increments, and $(\varepsilon_n)_{n\geq 0}$ is an $(\mathcal{F}_n)_{n\geq 0}$-adapted sequence of random variables with $\mathbb{E}[\varepsilon_{n+1}|\mathcal{F}_n] = 0$ almost surely.

For additional reading, one could refer to [Duf13], [Ben06], [Pem07]. The heuristic general idea is that, if F is regular enough, then $(X_n)_{n\geq 0}$ follows the flow of the differential equation $y' = F(y)$ (this is made rigorous by the concept of *pseudo-asymptotic trajectories* of [Ben06]), and this implies that $(X_n)_{n\geq 0}$ converges almost surely to the set of stable zeros of F.

4.3.1 Pólya urns are stochastic approximations

Lemma 4.20. *Let $(U(n))_{n\geq 0}$ be the d-colour Pólya urn of initial composition $\alpha = (\alpha_1, \ldots, \alpha_d)$ and replacement matrix $\mathcal{R} = (\mathcal{R}_{i,j})_{1\leq i,j\leq d}$. For every $n \geq 0$, set $T(n) = \|U(n)\|_1$ and $\hat{U}(n) = \frac{U(n)}{T(n)}$. Then, for all $n \geq 0$,*

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

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

$$S = \left\{(x_1, \dots, x_d) \in [0,1]^d : \sum_{i=1}^{d} x_i = 1\right\} \quad \text{and} \quad F(x) = \sum_{i=1}^{d} x_i(\mathcal{R}_i - \|\mathcal{R}_i\|_1 x),$$

for all $x \in S$.

Proof. Let $(\mathcal{F}_n)_{n\geq 0}$ be the natural filtration of $(U(n))_{n\geq 0}$. Recall that, $U(n+1) = U(n) + \mathcal{R}_{Y(n+1)}$, where $\mathbb{P}(Y(n+1) = i|\mathcal{F}_n) = \hat{U}_i(n)$, for every $1 \leq i \leq d$ (see section 3.1). Hence, for all $n \geq 0$,

$$\begin{aligned}\hat{U}(n+1) &= \frac{U(n) + \mathcal{R}_{Y(n+1)}}{T(n+1)} = \hat{U}(n) \cdot \frac{T(n)}{T(n+1)} + \frac{\mathcal{R}_{Y(n+1)}}{T(n+1)} \\ &= \hat{U}(n)\left(1 - \frac{\left\|\mathcal{R}_{Y(n+1)}\right\|_1}{T(n+1)}\right) + \frac{\mathcal{R}_{Y(n+1)}}{T(n+1)} \\ &= \hat{U}(n) + \frac{\mathcal{R}_{Y(n+1)} - \left\|\mathcal{R}_{Y(n+1)}\right\|_1 \hat{U}(n)}{T(n+1)}.\end{aligned}$$

Set

$$\Delta M_{n+1} = \mathcal{R}_{Y(n+1)} - \left\|\mathcal{R}_{Y(n+1)}\right\|_1 \hat{U}(n) - \mathbb{E}\Big[\mathcal{R}_{Y(n+1)} - \left\|\mathcal{R}_{Y(n+1)}\right\|_1 \hat{U}(n)\Big|\mathcal{F}_n\Big],$$

and notice that,

$$\mathbb{E}\Big[\mathcal{R}_{Y(n+1)} - \left\|\mathcal{R}_{Y(n+1)}\right\|_1 \hat{U}(n)\Big|\mathcal{F}_n\Big] = \sum_{i=1}^{d} \hat{U}_i(n)\big(\mathcal{R}_i - \|\mathcal{R}_i\|_1 \hat{U}(n)\big).$$

So, if we set $F : \mathbb{R}^d \to \mathbb{R}^d$ such that $F(x) = \sum_{i=1}^{d} x_i(\mathcal{R}_i - \|\mathcal{R}_i\|_1 x)$, and $\gamma_n = \frac{1}{T(n+1)}$, we get

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

□

Remark 4.21. If R is balanced, i.e. for every $1 \leq i \leq d$, $\|\mathcal{R}_i\|_1 = \mathcal{S} > 0$, then $T(n) = \|\alpha\|_1 + n\mathcal{S}$ for all $n \geq 0$, and $F(x) = x(\mathcal{R} - \mathcal{S} \cdot \mathrm{Id})$. In this case, we recover the result below.

Theorem 4.22 (Robbins-Monro, Theorem 1.4.26 in [Duf13])**.** *Assume that* $(X_n)_{n\geq 0}$ *is a stochastic approximation on* $S \subset \mathbb{R}^d$ *in the sense of Definition 4.19. Assume that there is a function* $\sigma^2 : S \to \mathbb{R}$ *such that*

(i) F *is continuous,*

(ii) $\mathbb{E}[(\Delta M_{n+1} + \varepsilon_{n+1})^2|\mathcal{F}_n] \leq \sigma^2(X_n)$ *for all* $n \geq 0$,

(iii) there exists a constant $K > 0$ *such that, for all* $x \in \mathbb{R}^d$,

$$\|F(x)\|_2^2 + \sigma^2(x) \leq K(1 + \|x\|_2^2),$$

(iv) there exists $x^* \in \mathbb{R}^d$ *such that* $F(x^*) = 0$ *and, for every* $x \in S \setminus \{x^*\}$,

$$\langle F(x), x - x^* \rangle < 0.$$

Then, $X_n \xrightarrow{n\to\infty} x^*$ *almost surely.*

Before proving this theorem, we show another result that will facilitate us in proving Robbins-Monro theorem.

Theorem 4.23 (Robbins-Siegmund, Theorem 1.3.12 in [Duf13])**.** *Suppose that* $(V_n)_{n\geq 0}$, $(\beta_n)_{n\geq 0}$, $(\chi_n)_{n\geq 0}$ *and* $(\eta_n)_{n\geq 0}$ *are four positive* $\mathcal{F}_n$*-adapted sequences, and that*

$$\mathbb{E}[V_{n+1}|\mathcal{F}_n] \leq V_n(1+\beta_n) + \chi_n - \eta_n.$$

Then, on $\Lambda = \Big\{ \sum_{n\geq 0} \beta_n < \infty \quad \text{and} \quad \sum_{n\geq 0} \chi_n < \infty \Big\}$, *almost surely* $(V_n)_{n\geq 0}$ *converges to a finite random variable* V *and the series* $\sum_{n\geq 0} \eta_n$ *converges.*

Proof. Let $\alpha_n = \big(\prod_{k=0}^n (1+\beta_k)\big)^{-1}$, $V_n' = \alpha_{n-1} V_n$, $\chi_n' = \alpha_n \chi_n$ and $\eta_n' = \alpha_n \eta_n$. Then, we have

$$\mathbb{E}[V_{n+1}'|\mathcal{F}_n] = \alpha_n \mathbb{E}[V_{n+1}|\mathcal{F}_n] \leq \alpha_n (V_n(1+\beta_n) + \chi_n - \eta_n) = V_n' + \chi_n' - \eta_n'.$$

Let $X_n = V_n' - \sum_{k=0}^{n-1} (\chi_k' - \eta_k')$, then notice that

$$\mathbb{E}[\Delta X_{n+1}|\mathcal{F}_n] = \mathbb{E}[V_{n+1}' - (V_n' + \chi_n' - \eta_n')|\mathcal{F}_n] \leq 0,$$

that means $(X_n)_{n\geq 0}$ is a supermartingale. Now, consider

$$T_a = \inf_{n\geq 0} \Big\{ \sum_{i=0}^n (\chi_i' - \eta_i') \geq a \Big\},$$

which is an $\mathcal{F}_n$-stopping time: indeed, by the nature of the sequences $(\chi_n)_{n\geq 0}$ and $(\eta_n)_{n\geq 0}$, we have

$$\{T_a \leq n\} = \bigcup_{k=0}^n \underbrace{\Big\{ \sum_{i=0}^k (\chi_i' - \eta_i') \geq a \Big\}}_{\in \mathcal{F}_k \subseteq \mathcal{F}_n} \in \mathcal{F}_n.$$

By the optional stopping theorem (Theorem 2.11), $(X_{n\wedge T_a})_{n\geq 0}$ is also a supermartingale, and since $n \wedge T_a - 1 \leq T_a$, by definition of T_a, we have $\sum_{i=0}^{n\wedge T_a - 1} (\chi_i' - \eta_i') < a$. Thus,

$$X_{n\wedge T_a} = V_{n\wedge T_a}' - \sum_{i=0}^{n\wedge T_a - 1} (\chi_i' - \eta_i') \geq -a.$$

Since the supermartingale is bounded from below, we can conclude that it converges almost surely to a finite random variable. So, on $\{T_a = \infty\}$, X_n converges almost surely to a finite limit. On $\Lambda = \Big\{ \sum_{n\geq 0} \beta_n < \infty \quad \text{and} \quad \sum_{n\geq 0} \chi_n < \infty \Big\}$,

$$\log\Big(\prod_{k=0}^n (1+\beta_k) \Big) = \sum_{k=0}^n \log(1+\beta_k) < \infty$$

as $n \to \infty$. This implies that $\alpha_n \to \alpha$ as $n \to \infty$, where α is a positive random variable, then $\sum_{n\geq 0} \chi_n < \infty$ implies that $\sum_{n\geq 0} \chi'_n < \infty$. On $\{T_a = \infty\} \cap \Lambda$, $(\eta_n)_{n\geq 0}$ being positive and the fact that

$$\sum_{k=0}^{n-1} \eta'_k \leq X_n + \sum_{k=0}^{n-1} \chi'_k < \infty,$$

implies that $\sum_{n\geq 0} \eta'_n < \infty$ and consequently $(V'_n)_{n\geq 0}$ converges. Thus, also $\sum_{n\geq 0} \eta_n$ and $(V_n)_{n\geq 0}$ converge. But since on Λ, $\sum_{n\geq 0} \chi'_n < \infty$ and

$$\Lambda = \bigcup_{N\geq 1} \big(\Lambda \cap \{T_N = \infty\}\big),$$

for every N, the theorem is proved. □

With the help of this theorem, we are ready to prove Robbins-Monro theorem:

Proof of Theorem 4.22. Consider $\|X_n - x^*\|_2^2$ and notice that,

$$\begin{aligned}\|X_{n+1} - x^*\|_2^2 &= \|X_n - x^* + \gamma_n(F(X_n) + \Delta M_{n+1} + \varepsilon_{n+1})\|_2^2 \\ &= \|X_n - x^*\|_2^2 + \gamma_n^2\|F(X_n) + \Delta M_{n+1} + \varepsilon_{n+1}\|_2^2 \\ &\quad + 2\gamma_n\langle X_n - x^*, F(X_n) + \Delta M_{n+1} + \varepsilon_{n+1}\rangle,\end{aligned}$$

and so,

$$\begin{aligned}\mathbb{E}\Big[\|X_n - x^*\|_2^2\Big|\mathcal{F}_n\Big] &= \|X_n - x^*\|_2^2 + \gamma_n^2\mathbb{E}\Big[\|F(X_n) + \Delta M_{n+1} + \varepsilon_{n+1}\|_2^2\Big|\mathcal{F}_n\Big] \\ &\quad + 2\gamma_n\langle X_n - x^*, F(X_n)\rangle.\end{aligned} \tag{4.9}$$

Analysing the second term on the right-hand side, we observe that, since $\mathbb{E}[\Delta M_{n+1} + \varepsilon_{n+1}|\mathcal{F}_n] = 0$,

$$\begin{aligned}\mathbb{E}\Big[\|F(X_n) + \Delta M_{n+1} + \varepsilon_{n+1}\|_2^2\Big|\mathcal{F}_n\Big] &= \|F(X_n)\|_2^2 + \mathbb{E}\Big[\|\Delta M_{n+1} + \varepsilon_{n+1}\|_2^2\Big|\mathcal{F}_n\Big] \\ &\leq \|F(X_n)\|_2^2 + \sigma^2(X_n) \\ &\leq K\big(1 + \|X_n\|_2^2\big).\end{aligned}$$

We also have, $\|F(x)\|_2^2 + \sigma^2(x) \leq K\big(1 + \|x\|_2^2\big) \leq K'\big(1 + \|x - x^*\|_2^2\big)$, for a finite positive constant K'. Hence, from equation (4.9) we get

$$\begin{aligned}\mathbb{E}\Big[\|X_{n+1} - x^*\|_2^2\Big|\mathcal{F}_n\Big] &\leq \|X_n - x^*\|_2^2 + \gamma_n^2K'\big(1 + \|X_n - x^*\|_2^2\big) + 2\gamma_n\langle X_n - x^*, F(X_n)\rangle \\ &= \|X_n - x^*\|_2^2\big(1 + \gamma_n^2K'\big) + \gamma_n^2K' + 2\gamma_n\langle X_n - x^*, F(X_n)\rangle.\end{aligned}$$

We can thus apply Theorem 4.23, and we get that $\|X_n - x^*\|_2^2 \xrightarrow{n\to\infty} V$ almost surely, where V is a finite random variable and the series $\sum_{n\geq 0} \gamma_n\langle X_n - x^*, F(X_n)\rangle$ converges almost surely.

Consider the situation where $\omega \in \Omega$ is outside of the negligible set, where convergence does not take place, and $\sum_{n\geq 0} \gamma_n < \infty$. If $V(\omega) \neq 0$, we can find two constants $0 < a < b < \infty$, such that for $n \geq n_0$, $a \leq \|X_n(\omega) - x^*\|_2 \leq b$. But, on the annulus $\{y : a \leq \|y - x^*\|_2 \leq b\}$, the continuous function $y \mapsto \langle F(y), y - x^* \rangle$ is bounded above by $-c$, for $c > 0$. This implies that $\sum_{n\geq 0} \gamma_n(\omega) < \infty$, which is a contradiction.
Hence, $V(\omega) = 0$, i.e. $X_n \xrightarrow{n\to\infty} x^*$ almost surely. □

We know show an application of this Robbins-Monro theorem to prove again Theorem 3.11.

Example 4.24. Consider a balanced Pólya urn $(U(n))_{n\geq 0}$ with initial composition $\alpha = (\alpha_1, \ldots, \alpha_d)$ and irreducible replacement matrix $\mathcal{R}$. By Lemma 4.20, $\hat{U}(n)$ is a stochastic approximation with $\gamma_n = \frac{1}{\|\alpha\|_1 + (n+1)\mathcal{S}}$, $F(x) = x(R - \mathcal{S} \cdot \mathrm{Id})$, $\Delta M_{n+1} = \mathcal{R}_{Y(n+1)} - \mathbb{E}[\mathcal{R}_{Y(n+1)} | \mathcal{F}_n]$, and $\varepsilon_{n+1} = 0$. We now show that this choice satisfies assumptions (i)-(iv) of Theorem 4.22:

- (i): F is linear, hence continuous.
- (ii): By the triangular inequality, we have $\|\Delta M_{n+1}\|_1 \leq 2\mathcal{S}$, so let $\sigma^2(x) = 2\mathcal{S}$ for all $x \in \mathbb{R}^d$.
- (iii) There exists $K > 0$ such that, for all $x \in \mathbb{R}^d$, $\|F(x)\|_2 = \|x(R - \mathcal{S} \cdot \mathrm{Id})\|_2 \leq K\|x\|_2$ (indeed, one can take K to be equal to the spectral radius of $R - \mathcal{S} \cdot \mathrm{Id}$).
- (iv) By Perron-Frobenius theorem (Theorem 3.10), there is a unique left-eigenvector $v \in S$ of R, with associated eigenvalue $\mathcal{S}$, therefore $F(v) = vR - \mathcal{S}v = 0$, and, for every $x \in S \setminus \{v\}$,

  $$\langle F(x), x - v \rangle = \langle x(R - \mathcal{S} \cdot \mathrm{Id}), x - v \rangle = \langle (x - v)(R - \mathcal{S} \cdot \mathrm{Id}), x - v \rangle < 0,$$

  since all the eigenvalues of $R - \mathcal{S} \cdot \mathrm{Id}$, except v, have negative real parts.

Hence, we can apply Theorem 4.22, showing that $\hat{U}(n) \xrightarrow{n\to\infty} v$ almost surely, and as we did in the proof of Theorem 3.11, it follows the same result of Theorem 3.11.

4.3.2 MVPPs are stochastic approximations

We now show that $(\hat{m}_n)_{n\geq 0}$ is a stochastic approximation. Note that $(\hat{m}_n)_{n\geq 0}$ takes values in $\mathcal{P}(E)$, the set of probability measures on E. Before doing that, we introduce the following 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 write

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

For any random probability measure μ, 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 4.25. *Let $(m_n)_{n\geq 0}$ be the stochastic approximation of initial composition m_0 and deterministic replacement kernel $R = (R_x)_{x\in E}$. Recall that, for every $i \geq 1$, $Y(i)$ denotes the colour of the ball drawn at time i. For all $n \geq 0$, set $\hat{m}_n = \frac{m_n}{m_n(E)}$. Then, for every $n \geq 0$,*

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

where $F : \mathcal{P}(E) \to \mathcal{P}(E)$ is defined 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. Recall that, for every $n \geq 0$, $m_{n+1} = m_n + R_{Y(n+1)}$. Thus,

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

Let $(\mathcal{F}_n)_{n\geq 0}$ be the canonical filtration generated by the $(Y(i))_{i\geq 1}$, then, since

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

we have

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

So, setting

$$\Delta M_{n+1} = R_{Y(n+1)} - R_{Y(n+1)}(E)\hat{m}_n - \mathbb{E}\big[R_{Y(n+1)} - R_{Y(n+1)}(E)\hat{m}_n\big|\mathcal{F}_n\big],$$

and defining $F : \mathcal{P}(E) \to \mathcal{P}(E)$, $\mu \mapsto \mu R - (\mu R)(E)\mu$, we get

$$\hat{m}_{n+1} = \hat{m}_n + \frac{1}{m_{n+1}(E)}\big(F(\hat{m}_n) + \Delta M_{n+1}\big).$$

□

Before presenting the main convergence result of non-balanced MMVPs, we would like to introduce some useful and necessary concepts to prove our result.

4.3.3 Quasi-stationary distributions

In this section, we provide a brief introduction to quasi-stationary distributions (QSD), based on [MV12], with definitions and results drawn from section 2 of the document. The next section will be based on [MV20], which presents the following definition for a QSD (also included in [Mai23]):

Definition 4.26 (QSD)**.** Let $(X_t)_{t\geq 0}$ 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 there exists some initial distribution $\pi \in \mathcal{P}(E)$ such that, for all Borel sets $A \subseteq E$,

$$\mathbb{P}_\pi(X_t \in A | X_t \neq \partial) \xrightarrow{t\to\infty} \nu(A),$$

where $\mathbb{P}_\pi$ is the law of X with initial distribution π.

However, to be precise, according to [MV12] (and other references, e.g. [CMSM13], [BCV21]), this is the definition of a *quasi-limiting distribution* (QLD). As we show in Proposition 4.29 below, a probability measure ν is a QSD for X if, and only if, it is also a QLD for X. To keep it less ambiguous, we will now provide the definitions of QSD and QLD, and in the next section, we will refer to a QSD in the sense of Definition 4.26. Let τ_∂ be the time when the process $(X_t)_{t\geq 0}$ is absorbed, i.e. the stopping time

$$\tau_\partial = \inf_{t\geq 0}\{X_t = \partial\}.$$

We consider the cases for which the process is almost surely absorbed, i.e. for every $x \in E$,

$$\mathbb{P}_x(\tau_\partial < \infty) = 1, \tag{4.10}$$

which can also be written as $\lim_{t\to\infty} \mathbb{P}_x(X_t \neq \partial) = 0$. We denote by $(P_t)_{t\geq 0}$ the transition kernel of the process X killed (or absorbed) at ∂, that is, for any positive x and f, continuous and bounded on E, we have

$$P_t f(x) = \mathbb{E}[f(X_t)\mathbb{1}_{\tau_\partial > t}].$$

Definition 4.27 (QSD'). Let $\nu \in \mathcal{P}(E)$. We say that ν is a quasi-stationary distribution for X if, for all $t \geq 0$ and all Borel sets $A \subseteq E$,

$$\mathbb{P}_\nu(X_t \in A | \tau_\partial > t) = \nu(A).$$

Definition 4.28 (QLD). Let $\nu \in \mathcal{P}(E)$. We say that ν is a quasi-limiting distribution for X, if there exists $\pi \in \mathcal{P}(E)$ such that, for every Borel set $A \subseteq E$,

$$\mathbb{P}_\pi(X_t \in A | \tau_\partial > t) \xrightarrow{t\to\infty} \nu(A).$$

Notice that, we have

$$\mathbb{P}_\pi(X_t \in A | \tau_\partial > t) = \frac{\mathbb{P}_\pi(X_t \in A, \tau_\partial > t)}{\mathbb{P}_\pi(\tau_\partial > t)} = \frac{\pi P_t \cdot \mathbb{1}_A}{\pi P_t \cdot \mathbb{1}_E}. \tag{4.11}$$

Proposition 4.29. *Let $\nu \in \mathcal{P}(E)$. Then, ν is a QLD for X if, and only if, it is a QSD for X.*

Proof. Suppose firstly that ν is a QSD for X. Then, it is straightforward that it is also a QLD for X starting with distribution ν.
On the other hand, assume ν is a QLD for X for an initial probability measure π on E. Thus, for any bounded and continuous function f on E, we have

$$\nu \cdot f = \lim_{t\to\infty} \mathbb{E}_\pi[f(X_t) | \tau_\partial > t]. \tag{4.12}$$

We claim that, letting $f(x) = \mathbb{P}_x(\tau_\partial > s)$, it holds that

$$\mathbb{P}_\nu(\tau_\partial > s) = \lim_{t\to\infty} \frac{\mathbb{P}_\pi(\tau_\partial > t + s)}{\mathbb{P}_\pi(\tau_\partial > t)}.$$

Indeed, notice that, on the event $\{\tau_\partial > t\}$, we have $\tau_\partial = t + \tau_\partial \circ \theta_t$. So,

$$\{\tau_\partial > t+s\} = \theta_t^{-1}\{\tau_\partial > s\} \cap \{\tau_\partial > t\},$$

or, in terms of indicator functions, $\mathbb{1}_{\tau_\partial > t+s} = \mathbb{1}_{\tau_\partial > t}\mathbb{1}_{\tau_\partial > s} \circ \theta_t$. Hence, by the simple Markov property (Theorem 2.17), for any $\mu \in \mathcal{P}(E)$,

$$\begin{aligned}\mathbb{P}_\mu\big(\tau_\partial > t+s\big) &= \mathbb{E}_\mu\big[\mathbb{1}_{\tau_\partial > t+s}\big] = \mathbb{E}_\mu\big[\mathbb{E}_\mu[\mathbb{1}_{\tau_\partial > s} \circ \theta_t | \mathcal{F}_t]\mathbb{1}_{\tau_\partial > t}\big] \\ &= \mathbb{E}_\mu\big[\mathbb{E}_{X_t}\big[\mathbb{1}_{\tau_\partial > s}\big]\mathbb{1}_{\tau_\partial > t}\big] = \mathbb{E}_\mu\big[\mathbb{P}_{X_t}\big(\tau_\partial > s\big)\mathbb{1}_{\tau_\partial > t}\big] \\ &= \mathbb{P}_\mu\big(\tau_\partial > t\big)\mathbb{E}_\mu\big[\mathbb{P}_{X_t}(\tau_\partial > s)\big|\tau_\partial > t\big].\end{aligned} \tag{4.13}$$

Hence, let $f(x) = \mathbb{P}_x(\tau_\partial > s)$ in relation (4.12), so we have

$$\mathbb{P}_\nu(\tau_\partial > s) = \lim_{t\to\infty} \mathbb{E}_\pi\big[\mathbb{P}_{X_t}(\tau_\partial > s)\big|\tau_\partial > t\big] = \lim_{t\to\infty} \frac{\mathbb{P}_\pi(\tau_\partial > t+s)}{\mathbb{P}_\pi(\tau_\partial > t)},$$

by equation (4.13).
Now, consider $f(x) = \mathbb{P}_x(X_s \in A, \tau_\partial > s)$, for $A \subseteq E$. Analogously as before, we get

$$\begin{aligned}\mathbb{P}_\nu(X_s \in A, \tau_\partial > s) &= \lim_{t\to\infty} \frac{\mathbb{P}_\pi(X_{t+s} \in A, \tau_\partial > t+s)}{\mathbb{P}_\pi(\tau_\partial > t)} \\ &= \lim_{t\to\infty} \frac{\mathbb{P}_\pi(X_{t+s} \in A, \tau_\partial > t+s)}{\mathbb{P}_\pi(\tau_\partial > t+s)} \cdot \frac{\mathbb{P}_\pi(\tau_\partial > t+s)}{\mathbb{P}_\pi(\tau_\partial > t)}.\end{aligned}$$

As $t \to \infty$, by definition of the QLD ν, the first term of the last expression converges to $\nu(A)$, and the second term converges to $\mathbb{P}_\pi(\tau_\partial > s)$. Therefore, this means that, for any Borel set $A \subseteq E$ and any $s > 0$, $\nu(A) = \mathbb{P}_\nu(X_s \in A | \tau_\partial > s)$, which concludes the proof. □

This proposition clarifies and reduces the ambiguity in the definitions of QSD presented earlier.

Proposition 4.30. *Let ν be a QSD for the process X with absorbing state ∂, satisfying condition (4.10). Then, there exists $\lambda_0 := \lambda_0(\nu) \in \mathbb{R}_{>0}$, such that*

$$\mathbb{P}_\nu(\tau_\partial > t) = e^{-\lambda_0 t}.$$

Proof. We have, as we derived in relation (4.13),

$$\mathbb{P}_\mu\big(\tau_\partial > t+s\big) = \mathbb{P}_\mu\big(\tau_\partial > t\big)\mathbb{E}_\mu\big[\mathbb{P}_{X_t}(\tau_\partial > s)\big|\tau_\partial > t\big].$$

By definition of QSD, we get $\mathbb{E}_\nu[\mathbb{P}_{X_t}(\tau_\partial > s)|\tau_\partial > t] = \mathbb{P}_\nu(\tau_\partial > s)$. Thus, for any $s, t > 0$,

$$\mathbb{P}_\nu\big(\tau_\partial > t+s\big) = \mathbb{P}_\nu\big(\tau_\partial > t\big)\mathbb{P}_\nu\big(\tau_\partial > s\big).$$

Let $g(t) = \mathbb{P}_\nu(\tau_\partial > t)$ (this function is often referred to as *survival function*, see e.g. [KK96]). We have $g(0) = 1$, and by condition (4.10), we deduce that $g(t) \to 0$ as $t \to \infty$. Notice that $g(t) = e^{-\lambda_0 t}$ is a good candidate for a survival function: indeed, it satisfies the linear differential equation $g'(t) = -\lambda_0 g(t)$ with initial condition $g(0) = 1$. This shows that there exists a real positive λ_0 such that $\mathbb{P}_\nu(\tau_\partial > t) = e^{-\lambda_0 t}$. □

We can now derive a nice tool, often used in the literature studying QSDs.

Proposition 4.31. *Let $\nu \in \mathcal{P}(E)$ and let Q be the jump kernel of the sub-Markovian transition kernel $(P_t)_{t\geq 0}$ on E. If ν is a QSD, then there is a $\lambda_0 > 0$ such that, for every function f continuous and bounded,*

$$\nu Q \cdot f = -\lambda_0 \nu \cdot f.$$

Remark 4.32. The implication in the other direction also holds, but for our purposes we focus only on this one.

Proof. By definition of QSD, for every Borel set $A \subseteq E$, we have (see relation (4.11)),

$$\nu(A) = \frac{\nu P_t \cdot \mathbb{1}_A}{\nu P_t \cdot \mathbb{1}_E}.$$

By Proposition 4.30, there exists $\lambda_0 \in \mathbb{R}_{>0}$ such that, for any $t \geq 0$,

$$\nu P_t \cdot \mathbb{1}_E = \mathbb{P}_\nu(\tau_\partial > t) = e^{-\lambda_0 t},$$

and, consequently

$$\nu P_t \cdot \mathbb{1}_A = e^{-\lambda_0 t}\nu(A),$$

which is equivalent to $\nu P_t = e^{-\lambda_0 t}\nu$. By Kolmogorov's forward equation (equation 2.1), we have, for any function f continuous and bounded,

$$\left|\frac{\mathrm{d}}{\mathrm{d}t}P_t f(x)\right| = |P_t Q \cdot f(x)| \leq \|Q \cdot f\|_\infty < \infty. \tag{4.14}$$

In particular, we can differentiate $\nu P_t \cdot f = \int_E P_t \cdot f(x)\,\mathrm{d}\nu(x)$ under the integral sign, so for the left-hand side, we get

$$\frac{\mathrm{d}}{\mathrm{d}t}\nu P_t \cdot f = \frac{\mathrm{d}}{\mathrm{d}t}e^{-\lambda_0 t}\nu \cdot f = -\lambda_0 e^{-\lambda_0 t}\nu \cdot f.$$

For the right-hand side, we get, by relation (4.14),

$$\frac{\mathrm{d}}{\mathrm{d}t}\left[\int_E P_t \cdot f(x)\,\mathrm{d}\nu(x)\right] = \int_E \frac{\mathrm{d}}{\mathrm{d}t}P_t \cdot f(x)\,\mathrm{d}\nu(x) = \int_E P_t Q \cdot f(x)\,\mathrm{d}\nu(x)\,.$$

Since these hold for any $t \geq 0$, in particular for $t = 0$, we get

$$\nu Q \cdot f = -\lambda_0 \nu \cdot f,$$

as desired. □

With these new notions, we are prepared to state the main convergence result for non-balanced MVPPs.

4.3.4 Almost sure convergence of non-balanced MVPPs

In what follows, we denote by R the common expectation of the sequence of i.i.d. replacement kernels $(R^{(n)})_{n\geq 0}$, according to Definition 4.1.

Theorem 4.33. *Let $(m_n)_{n\geq 0}$ be the MVPP of initial composition m_0 and replacement kernel R. We assume that:*

(a) For all $x \in E$, $R_x(E) \leq 1$, and there exists $\rho \in \mathcal{P}(\mathbb{R})$ with positive mean such that, for every $x \in E$, $\mathcal{L}(R_x^{(i)}(E))$ stochastically dominates ρ, i.e. for a random variable Z with law ρ, we have $\mathbb{P}(R_x^{(i)}(E) > t) \geq \mathbb{P}(Z > t)$, for all real t. In particular, setting $c = \int_0^\infty x \, \mathrm{d}\rho(x)$,

$$0 < c \leq R_x(E) \leq 1.$$

(b) There exists a locally bounded function $V : E \to [1, \infty)$ such that:

(i) for every $L \geq 0$, $\{x \in E : V(x) \leq L\}$ is relatively compact,

(ii) there exists $\theta \in (0, c)$ and $K \geq 0$ such that

$$R_x \cdot V \leq \theta V(x) + K \quad (\forall x \in E),$$

(iii) there exist $r > 1, p > \frac{\log(2)}{\log(\theta/c)} \vee 2, A > 0$ such that

$$\mathbb{E}[R_x^{(1)}(E)^r] \vee \mathbb{E}[R_x^{(1)}(E)^p] \leq AV(x) \quad (\forall x \in E).$$

(c) The continuous-time jump process of sub-Markovian jump kernel $R - \mathrm{Id}$ admits a QSD ν (in the sense of Definition 4.26). Assume also that the convergence of $\mathbb{P}_\pi(X_t \in \cdot | X_t \neq \partial)$ holds uniformly with respect to the total variation norm on $\{\pi \in \mathcal{P}(E) : \pi \cdot V^{1/q} \leq C\}$, for each $C > 0$, where $q = \frac{p}{p-1}$.

(d) For every function $f : E \to \mathbb{R}$ bounded and continuous, $x \in E \mapsto R_x f$ is continuous.

Further, if $m_0 \cdot V < \infty$, then $\frac{m_n}{n} \to \nu R$ almost surely as $n \to \infty$, with respect to the weak topology. Moreover, $\sup_n \{m_n \cdot V^{1/q}/n\} < \infty$ almost surely. Furthermore, if $\nu R(E) > 0$, then, with respect to the weak topology,

$$\hat{m}_n \xrightarrow{n\to\infty} \nu \quad \text{a.s.}$$

In order to prove Theorem 4.33, we first want to prove the almost sure convergence of an auxiliary sequence of random distributions $(\eta_n)_{n\geq 0}$, which we will define soon. However, this convergence holds under an alternative of assumption (b) of Theorem 4.33. As discussed in section 1.4 of [MV20], this alternative is needed in the case when we assume that the measures $(R_x)_{x\in E}$ are not necessarily positive, but we consider these measures as signed kernels. In this case, we recover the following assumptions:

(b') There exist a locally bounded function $V : E \to [1, \infty)$, and some constants $\theta \in (0, c), K \geq 0, p > 2, q = \frac{p}{p-1}$, $q' > q$, $r > 1$, $A \geq 1$ and $B \geq 1$ such that:

(i) for every $L \geq 0$, $\{x \in E : V(x) \leq L\}$ is relatively compact,

(ii) $R_x \cdot V \leq \theta V(x) + K \quad \text{and} \quad R_x \cdot V^{1/q} \leq \theta V^{1/q}(x) + K \quad (\forall x \in E)$,

(iii) for every function $f : E \to \mathbb{R}$ continuous, bounded by 1, and every $x \in E$,

$$|R_x \cdot f|^{q'} \vee \mathbb{E}\Big[\Big|R_x^{(i)} \cdot f - R_x \cdot f\Big|^r\Big] \vee \mathbb{E}\Big[\Big|R_x^{(i)} \cdot f - R_x \cdot f\Big|^p\Big] \leq AV(x),$$

(iv) $\big|R_x \cdot V^{1/q}\big|^q \vee |R_x \cdot V| \vee \mathbb{E}\Big[\Big|R_x^{(i)} \cdot V^{1/q} - R_x \cdot V^{1/q}\Big|^r\Big] \leq BV(x).$

(e) For all $n \geq 0$, m_n is almost surely a positive measure.

Assuming that assumptions (a),(c) and (d) are also met, the conclusions of Theorem 4.33 hold true. As we show in the following, the set of "new" assumptions is actually implied by the original assumptions of the theorem, thus we prove the latter in the more general result.

Lemma 4.34. *Assumptions (a)-(d) imply assumptions (a),(b'),(c),(d) and (e).*

Proof. Assumption (e) follows by the assumption that $(R_x)_{x\in E}$ is positive.
Now, fix $q = \frac{p}{p-1}$; by Hölder's inequality (Theorem 2.1) ($q \geq 1$), for every $x \in E$, we have

$$R_x \cdot V^{1/q} \leq R_x(E)^{1/p}(R_x \cdot V)^{1/q},$$

and hence, also by assumption (b),

$$\big(R_x \cdot V^{1/q}\big)^q \leq R_x(E)^{q/p} R_x \cdot V \leq \theta V(x) + K.$$

By concavity, for $u \geq 0$ and $\alpha \leq 1$, it holds $(1+u)^\alpha \leq 1 + u^\alpha$, thus,

$$R_x \cdot V^{1/q} \leq (\theta V(x) + K)^{1/q} \leq \theta^{1/q} V^{1/q}(x) + K^{1/q}.$$

Now, notice that

$$\frac{1}{q}\log(\theta) = \left(1 - \frac{1}{p}\right)\log(\theta) < \left(1 - \frac{\log(\theta/c)}{\log(\theta)}\right)\log(\theta) = \log(c),$$

since $p > \frac{\log(\theta)}{\log(\theta/c)}$, and hence, $\big(1 - \frac{1}{p}\big) > \Big(1 - \frac{\log(\theta/c)}{\log(\theta)}\Big)$. Therefore, $\theta^{1/q} < c$, which proves (b'-ii).
Now we prove (b'-iii): let $q' \coloneqq p > q > 1$, then using Jensen's inequality (Theorem 2.2),

$$|R_x \cdot f|^{q'} = \mathbb{E}\Big[\Big|R_x^{(1)} \cdot f\Big|\Big]^{q'} \leq \mathbb{E}\Big[\Big|R_x^{(1)} \cdot f\Big|^{q'}\Big] \leq \mathbb{E}\big[R_x^{(1)}(E)^{q'}\big],$$

where the last inequality holds for every continuous function bounded by 1. Then, $\mathbb{E}\big[R_x^{(1)}(E)^{q'}\big] \leq AV(x)$ by assumption (b-iii). Analogously, for $1 < r' \leq r$, we have, using that, for any convex function φ, $\varphi\big(\frac{x+y}{2}\big) \leq \frac{\varphi(x)+\varphi(y)}{2}$, and Jensen's inequality,

$$\begin{aligned}\mathbb{E}\Big[\Big|R_x^{(1)} \cdot f - R_x \cdot f\Big|^p\Big] &\leq 2^{r'-1}\mathbb{E}\Big[\Big|R_x^{(1)} \cdot f\Big|^{r'} + |R_x \cdot f|^{r'}\Big] = 2^{r'}\mathbb{E}\Big[\Big|R_x^{(1)} \cdot f\Big|^{r'}\Big] \\ &\leq 2^{r'}\mathbb{E}[R_x^{(1)}(E)] \leq 2^{r'} AV(x),\end{aligned}$$

where the last inequality is due to (b-iii); in a similar way it is shown that

$$\mathbb{E}\Big[\Big|R_x^{(1)} \cdot f - R_x \cdot f\Big|^p\Big] \leq AV(x).$$

To prove (b'-iv), we start by noting that, since $V(x) \geq 1$, and since R_x is non-negative, using Hölder's inequality, we get

$$\begin{aligned}\Big|R_x \cdot V^{1/q}\Big|^q = (R_x \cdot V^{1/q})^q &\leq R_x(E)^{q/p} R_x \cdot V \\ &\leq R_x \cdot V \leq \theta V(x) + K \leq (\theta + K)V(x),\end{aligned}$$

where we also used assumption (b-ii). Now, arguing as before,

$$\begin{aligned}\mathbb{E}\Big[\Big|R_x^{(1)} \cdot V^{1/q} - R_x \cdot V^{1/q}\Big|^{r'}\Big] &\leq 2^{r'-1}\mathbb{E}\Big[\Big|R_x^{(1)} \cdot V^{1/q}\Big|^{r'} + \Big|R_x \cdot V^{1/q}\Big|^{r'}\Big] \\ &\leq 2^{r'}\mathbb{E}\big[(R_x^{(1)} \cdot V^{1/q})^{r'}\big].\end{aligned}$$

By Hölder's inequality, it holds

$$R_x^{(1)} \cdot V^{1/q} \leq (R_x^{(1)} \cdot V)^{1/q} R_x^{(1)}(E)^{1/p}.$$

Set $\omega = \frac{q}{r'}$, then, by Hölder's inequality,

$$\begin{aligned}\mathbb{E}\big[(R_x^{(1)} \cdot V^{1/q})^{r'}\big] &\leq \mathbb{E}\big[(R_x^{(1)} \cdot V)^{\frac{\omega r'}{q}}\big]^{1/\omega} \mathbb{E}\big[R_x^{(1)}(E)^{\frac{r'\omega}{p(\omega-1)}}\big]^{\frac{\omega-1}{\omega}} \\ &\leq \mathbb{E}\big[R_x^{(1)} \cdot V\big]^{1/\omega} \mathbb{E}\big[1 + R_x^{(1)}(E)^p\big]^{\frac{\omega-1}{\omega}},\end{aligned}$$

where the last inequality follows by the fact, that $\alpha \mapsto c^\alpha$ is monotonically decreasing if $c \in (0,1)$, and since $\frac{r'\omega}{p(\omega-1)} = \frac{r'(q-1)}{q-r'} \leq p$, since $q = \frac{p}{p-1}$, for r' small enough in $(1, r]$. By (b-ii), $\mathbb{E}\big[R_x^{(1)} \cdot V\big] = R_x \cdot V \leq (\theta + K)V(x)$, and by (b-iii), $\mathbb{E}\big[R_x^{(1)}(E)^p\big] \leq AV(x)$. Therefore, we get

$$\mathbb{E}\big[(R_x^{(1)} \cdot V^{1/q})^{r'}\big] \leq (\theta + K + 1 + A)V(x),$$

where we used the fact that $\frac{1}{\omega} + \frac{\omega-1}{\omega} = 1$.
This concludes the proof. □

Proposition 4.35. *Let $\eta_0 = 0$ and*

$$\eta_n = \eta_{n-1} + \delta_{Y(n+1)} = \sum_{i=1}^{n} \delta_{Y(i)}.$$

Set $\hat{\eta}_0 = 0$ and, for every $n \geq 1$, $\hat{\eta}_n = \frac{\eta_n}{\eta_n(E)} = \frac{\eta_n}{n}$. Suppose that assumption (a), (b'), (c), (d), (e) hold, then

$$\hat{\eta}_n \xrightarrow{n\to\infty} \nu \quad \text{a.s.}$$

Proof. The complete proof can be found in [MV20] (see Proposition 7). Here, we provide a sketch of the proof and cite some of the key results used.

Consider the dynamical system defined by the following equation, for all bounded and continuous functions $f : E \to \mathbb{R}$:

$$\frac{\mathrm{d}\mu_t \cdot f}{\mathrm{d}t} = \mu_t R \cdot f - \mu_t R(E)\mu_t \cdot f, \tag{4.15}$$

where the family $(\mu_t)_{t\geq 0}$ does not depend on f.
To show the almost sure convergence of $\hat{\eta}_n$ to ν, a linearization of it is shown to be a *pseudo-asymptotic trajectory* (refer to section 3 of [Ben06]) of the semi-flow induced by the dynamical system (4.15). We will not go into these topics in detail, which is why we are not including the complete proof. However, we cite the steps necessary to prove the proposition:

1. The studied stochastic algorithm is formulated, i.e. it is shown that $\hat{\eta}_n$ is a stochastic approximation in the sense of Definition 4.19.
2. It is proved, that the expectation of V with respect to the measure valued process remains bounded (see Lemma 4.37 below).
3. Almost sure convergence of the quantity introduced in Proposition 4.1 of [Ben06] is established to control the error term between the dynamical system (4.15) and its linearized counterpart.
4. It is shown, that the sequence $(\hat{\eta}_n)_{n\geq 0}$ is relatively compact for the topology of weak convergence on $\mathcal{P}(E)$.
5. Existence, uniqueness, and continuity properties of the flow induced by this system are established and proved.

□

We limit ourselves to proving the following two results, which will be needed for the proof of Theorem 4.33: the first is not included in the list of steps above, but was needed to prove them, and the second is point 2 in the list above.

Lemma 4.36. *Fix $c' \in (\theta, c)$. For all $k \geq 1$, let*

$$\sigma_k = \inf_{n\geq k} \{m_n(E) < c'n\}.$$

Then,

$$\mathbb{P}\left(\bigcup_{k\geq 1}\{\sigma_k = \infty\}\right) = 1.$$

To prove this lemma, we also use a technique called *coupling*: coupling is an important technique in probability theory, used to analyse the interactions and behaviours of random variables or stochastic processes by defining them within a common probability space. This method involves the construction of two or more random variables simultaneously in such a way that they can be directly compared or manipulated, facilitating the study of their convergence, dependence, and other key properties. For a deeper insight into the topic, one could, for example, refer to [Lin02] (a more formal introduction is given in Chapter 1).

Proof. Recall that,

$$m_n(E) = m_0(E) + \sum_{i=1}^{n} R^{(i)}_{Y(i)}(E).$$

By assumption (a) and by coupling, conditionally on $(Y(i))_{i\geq 1}$, the independence of the random variables $R^{(i)}_{Y(i)}(E)$ implies that there exists a sequence of random variables $(Z_i)_{i\geq 1}$ with law ρ such that, conditionally on the $Y(1), Y(2), \dots$, we have $R^{(i)}_{Y(i)}(E) \geq Z_i$ for all $i \geq 1$. Then, by the law of large numbers (Theorem 2.6), we get

$$\liminf_{n\to\infty} \frac{m_n(E)}{n} \geq \liminf_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} Z_i = c' \quad \text{a.s.},$$

which proves the lemma. □

Lemma 4.37. *For every $k \geq 1$, there is a constant $C_k > 0$ such that*

$$\sup_{n\geq 1} \mathbb{E}\left[\frac{\eta_{n\wedge\sigma_k}\cdot V}{n}\right] \vee \sup_{n\geq 1} \mathbb{E}\left[\frac{m_{n\wedge\sigma_k}\cdot V}{n}\right] \vee \sup_{n\geq 1} \mathbb{E}\big[V(Y(n+1))\mathbb{1}_{n<\sigma_k}\big] \leq C_k.$$

Proof. Fix $n \geq k+1$, we have

$$\begin{aligned}
\mathbb{E}\left[\frac{\eta_{(n+1)\wedge\sigma_k}\cdot V}{n+1}\right] &= \mathbb{E}\left[\frac{\eta_n\cdot V + \delta_{Y(n+1)}\cdot V}{n+1}\mathbb{1}_{n<\sigma_k}\right] + \mathbb{E}\left[\frac{\eta_{\sigma_k}\cdot V}{n+1}\mathbb{1}_{n\geq\sigma_k}\right] \\
&= \mathbb{E}\left[\frac{\eta_{n\wedge\sigma_k}\cdot V}{n+1}\right] + \frac{\mathbb{E}[V(Y(n+1))\mathbb{1}_{n<\sigma_k}]}{n+1} \qquad (4.16)\\
&= \Big(1-\frac{1}{n+1}\Big)\mathbb{E}\left[\frac{\eta_{n\wedge\sigma_k}\cdot V}{n}\right] + \frac{\mathbb{E}[V(Y(n+1))\mathbb{1}_{n<\sigma_k}]}{n+1}.
\end{aligned}$$

By definition of σ_k, for all $n \in \{k+1, \dots, \sigma_k - 1\}$, we have $m_n(E) \geq c'n$. Recalling that we assumed m_n to be almost surely a positive measure, we have that, since $Y(n+1)$ is distributed as $\frac{m_n}{m_n(E)}$, it holds, for every $n \geq k+1$,

$$\begin{aligned}
\mathbb{E}\big[V(Y(n+1))\mathbb{1}_{n<\sigma_k}\big] &= \mathbb{E}\left[\frac{m_n\cdot V}{m_n(E)}\mathbb{1}_{n<\sigma_k}\right] \\
&\leq \frac{1}{c'n}\mathbb{E}\big[m_{n\wedge\sigma_k}\cdot V\big] \\
&= \frac{1}{c'n}\mathbb{E}\left[m_0\cdot V + \sum_{i=1}^{n\wedge\sigma_k} R^{(i)}_{Y(i)}\cdot V\right] \\
&= \frac{1}{c'n}\mathbb{E}\left[m_0\cdot V + \sum_{i=1}^{n} R^{(i)}_{Y(i)}\cdot V\mathbb{1}_{i\leq\sigma_k}\right] \\
&= \frac{1}{c'n}\mathbb{E}\left[m_0\cdot V + \sum_{i=1}^{n} R_{Y(i)}\cdot V\mathbb{1}_{i\leq\sigma_k}\right].
\end{aligned}$$

Notice that the last equality is obtained by conditioning on $\mathcal{F}_{i-1}$ and $Y(i)$: $\mathbb{1}_{i\leq\sigma_k}$ is measurable with respect to $\mathcal{F}_i \cup \sigma(Y(i))$, then, since $\mathbb{1}_{i\leq\sigma_k} \leq \mathbb{1}_{i-1\leq\sigma_k}$, it is also $\mathcal{F}_{i-1}$-measurable, and $R^{(i)}$ is independent of $\mathcal{F}_i \cup \sigma(Y(i))$.

Hence, by assumption (b'-ii),

$$\begin{aligned}\mathbb{E}\big[V(Y(n+1))\mathbb{1}_{n<\sigma_k}\big] &\leq \frac{1}{c'n}\mathbb{E}\big[m_{n\wedge\sigma_k}\cdot V\big] = \frac{m_0\cdot V+\mathbb{E}[\eta_{n\wedge\sigma_k}R\cdot V]}{c'n}\\ &\leq \frac{m_0\cdot V+nK+\theta\mathbb{E}[\eta_{n\wedge\sigma_k}\cdot V]}{c'n}\\ &= \frac{m_0\cdot V+nK}{c'n}+\frac{\theta}{c'}\mathbb{E}\left[\frac{\eta_{n\wedge\sigma_k}\cdot V}{n}\right].\end{aligned} \tag{4.17}$$

So, by equation (4.16),

$$\mathbb{E}\left[\frac{\eta_{(n+1)\wedge\sigma_k}\cdot V}{n+1}\right] \leq \left(1-\frac{1-\theta/c'}{n+1}\right)\mathbb{E}\left[\frac{\eta_{n\wedge\sigma_k}\cdot V}{n}\right]+\frac{m_0\cdot V+nK}{(n+1)c'n}.$$

Now, since $\frac{\eta_{n\wedge\sigma_k}}{n}\leq\hat{\eta}_{n\wedge\sigma_k}$,

$$\mathbb{E}[\hat{\eta}_{n\wedge\sigma_k}\cdot V] = \frac{1}{n\wedge\sigma_k}\mathbb{E}[\eta_{n\wedge\sigma_k}\cdot V] = \frac{1}{n\wedge\sigma_k}\mathbb{E}\left[\sum_{i=1}^{n\wedge\sigma_k}V(Y(i))\right]<\infty,$$

by definition of V. Then, since we assumed $m_o\cdot V<\infty$, and since $\theta<c'<c<1$, it follows that $\mathbb{E}\big[\frac{\eta_{n\wedge\sigma_k}\cdot V}{n}\big]$ is uniformly bounded in n. By the inequalities (4.17), we conclude that $\mathbb{E}\big[\frac{m_{n\wedge\sigma_k}\cdot V}{n}\big]$ and $\mathbb{E}[V(Y(n+1))\mathbb{1}_{n<\sigma_k}]$ are also uniformly bounded in n. □

We also highlight an important expression derived in the proof of point 4 (which is Lemma 6 in [MV20]), that will be crucial for our main result: it holds, for every $n\geq 0$, for two finite random variables M and $\hat{K}\geq K$, that

$$\hat{\eta}_n\cdot V^{1/q}\leq 2M+\frac{1+c-\theta}{c-\theta}\hat{K}, \tag{4.18}$$

i.e., what we are interested in, $\hat{\eta}_n\cdot V^{1/q}$ is almost surely uniformly bounded. With these results, we proceed to prove Theorem 4.33:

Proof of Theorem 4.33. Let $\eta_0=0$ and $\eta_n=\sum_{i=1}^n\delta_{Y(i)}$. Recall that

$$m_n = m_0+\sum_{i=1}^n R^{(i)}_{Y(i)} = m_0+\sum_{i=1}^n\delta_{Y(i)}R^{(i)},$$

where $Y(n+1)$ is of distribution $\frac{m_n}{m_n(E)}$. Set $\hat{\eta}_0=0$ and $\hat{\eta}_n=\frac{\eta_n}{n}$, for every $n\geq 1$. Fix $c'\in(\theta,c)$ and, for all $k\geq 1$, let

$$\sigma_k=\inf_{n\geq k}\{m_n(E)<c'n\}.$$

For every $n\geq 0$ and any bounded continuous function $f:E\to\mathbb{R}$, set

$$\Psi_n = m_{n\wedge\sigma_k}\cdot f-\eta_{n\wedge\sigma_k}R\cdot f.$$

Notice that,

$$\begin{aligned}\mathbb{E}[\Psi_{n+1}|\mathcal{F}_n] &= \mathbb{E}\big[m_{(n+1)\wedge\sigma_k}\cdot f - \eta_{n\wedge\sigma_k}R\cdot f\big|\mathcal{F}_n\big]\\ &= m_{n\wedge\sigma_k}\cdot f - \eta_{n\wedge\sigma_k}R\cdot f + \mathbb{E}\Big[\mathbb{1}_{n<\sigma_k}\Big(R^{(n+1)}_{Y(n+1)}\cdot f - R_{Y(n+1)}\cdot f\Big)\Big|\mathcal{F}_n\Big],\end{aligned}$$

where the last term of the summation vanishes, since R is the common expectation of the $R^{(n)}$'s. Hence, $(\Psi_n)_{n\geq 0}$ is a martingale with

$$\Psi_n = m_0\cdot f + \sum_{i=1}^{n\wedge\sigma_k}\Big(R^{(i)}_{Y(i)}\cdot f - R_{Y(i)}\cdot f\Big).$$

We take now two results from the literature, that will help us throughout the theorem:

Theorem 4.38 (Theorem 1.3.17 of [Duf13])**.** *Let α and β be such that $0<\alpha<2\beta$. Let $(M_n)_{n\geq 0}$ be a martingale in L^2. Suppose that $n^{-\alpha}\mathbb{E}[\langle M\rangle_n]$ is bounded for every $n\geq 1$. Then,*

$$\frac{M_n}{n^\beta}\xrightarrow{n\to\infty}0 \quad \text{a.s.}$$

An immediate adaptation of this theorem tells us that, if $n^{-1}\mathbb{E}[|\Psi|^p]$ is bounded for all $n\geq 1$, then $\frac{\Psi_n}{n}\to 0$ almost surely as $n\to\infty$, for $p>1$. Further, we also have,

Lemma 4.39 (Lemma 1 of [Cha69])**.** *If $\mathbb{E}[M_j|M_1+\cdots+M_{j-1}]=0$ (in particular, if M_j is a martingale difference sequence), for $2\leq j\leq n$ and $M_j\in L^p$, for $1\leq p\leq 2$, then*

$$\mathbb{E}\Big[|M_1+\cdots+M_n|^p\Big] = \int|M_1+\cdots+M_n|^p \leq \alpha\sum_{j=1}^n\int|M_j|^p,$$

where $\alpha\leq 2^{2-p}$.

Using this result, we get

$$\frac{\mathbb{E}\big[|\Psi_n|^p\big]}{n}\leq\frac{2}{n}\sum_{j=0}^n\mathbb{E}\big[|\Psi_j|^p\big] = \frac{2(m_0\cdot f)^p}{n}+\frac{2}{n}\sum_{i=1}^n\mathbb{E}\Big[\Big|R^{(i)}_{Y(i)}\cdot f - R_{Y(i)}\cdot f\Big|^p\mathbb{1}_{i\leq\sigma_k}\Big], \tag{4.19}$$

where

$$\begin{aligned}\mathbb{E}\Big[\Big|R^{(i)}_{Y(i)}\cdot f - R_{Y(i)}\cdot f\Big|^p\mathbb{1}_{i\leq\sigma_k}\Big] &= \mathbb{E}\Big[\mathbb{E}\Big[\Big|R^{(i)}_{Y(i)}\cdot f - R_{Y(i)}\cdot f\Big|^p\mathbb{1}_{i\leq\sigma_k}\Big|\mathcal{F}_{i-1}\Big]\Big]\\ &= \mathbb{E}\Big[\mathbb{E}\Big[\Big|R^{(i)}_{Y(i)}\cdot f - R_{Y(i)}\cdot f\Big|^p\Big]\mathbb{1}_{i\leq\sigma_k}\Big],\end{aligned}$$

because $\mathbb{1}_{i\leq\sigma_k}\leq\mathbb{1}_{i-1\leq\sigma_k}$, and so $\mathbb{1}_{i\leq\sigma_k}$ is $\mathcal{F}_{i-1}$-measurable. Thus, by assumption (b'-iii),

$$\mathbb{E}\Big[\mathbb{E}\Big[\Big|R^{(i)}_{Y(i)}\cdot f - R_{Y(i)}\cdot f\Big|^p\Big]\mathbb{1}_{i\leq\sigma_k}\Big]\leq\|f\|^p_\infty A\mathbb{E}\big[V(Y(i))\mathbb{1}_{i\leq\sigma_k}\big].$$

Bringing everything together, we have

$$\frac{\mathbb{E}\big[|\Psi_n|^p\big]}{n} \leq \frac{2(m_0 \cdot f)^p}{n} + \frac{2\|f\|_\infty^p}{n} A \sum_{i=1}^{n} \mathbb{E}\big[V(Y(i))\mathbb{1}_{i\leq\sigma_k}\big].$$

By Lemma 4.37, it follows that $\frac{\mathbb{E}[|\Psi_n|^p]}{n}$ is bounded for all $n \geq 1$, since, by assumption, $m_o \cdot V < \infty$, which entails $m_0(E) < \infty$. Thus, the sequence $(n^{-1}\mathbb{E}[|\Psi_n|^p])_{n\geq 1}$ converges to 0 almost surely. This is true for any $k \geq 1$, and, by Lemma 4.36, we deduce that, $m_n \cdot f = \eta_n R \cdot f + o(n)$ almost surely. By Proposition 4.35 and assumption (d), it follows that

$$\eta_n R \cdot \frac{f}{n} \xrightarrow{n\to\infty} \nu R \cdot f \quad \text{a.s. and} \quad \frac{\eta_n R(E)}{n} \xrightarrow{n\to\infty} \nu R(E).$$

By Proposition 4.31 and assumption (c), we have that, there exists a constant λ_0 such that $\nu(R - \mathrm{Id}) = -\lambda_0 \nu$. In other words, the quasi-stationary distribution ν is a left eigenfunction for R, with associated eigenvalue $\lambda \coloneqq 1 - \lambda_0 \in (0, 1]$, and hence we have the convergence

$$\hat{m}_n = \frac{m_n}{m_n(E)} \xrightarrow{n\to\infty} \frac{\nu R}{\nu R(E)} = \frac{\lambda\nu}{\lambda\nu(E)} = \nu \quad \text{a.s.}$$

Now we want to show that $m \cdot V^{1/q}/n$ is bounded almost surely. For every $n \geq 0$,

$$m_n \cdot V^{1/q} = m_0 \cdot V^{1/q} + \sum_{i=1}^{n} R^{(i)}_{Y(i)} \cdot V^{1/q}.$$

As before, let

$$\Phi_n = m_0 \cdot V^{1/q} + \sum_{i=1}^{n\wedge\sigma_k} (R^{(i)}_{Y(i)} \cdot V^{1/q} - R_{Y(i)} \cdot V^{1/q}).$$

With the same argument as before, we get that $(\Phi_n)_{n\geq 0}$ is a martingale and that $\frac{\Phi_n}{n} \to 0$ almost surely as $n \to \infty$. Hence,

$$\frac{m_n \cdot V^{1/q}}{n} = \frac{1}{n} \sum_{i=1}^{n} R_{Y(i)} \cdot V^{1/q} + o(1) = \hat{\eta}_n R \cdot V^{1/q} + o(1).$$

Assumption (b'-iv) implies that

$$\left|\hat{\eta}_n R \cdot V^{1/q}\right| \leq B^{1/q} \hat{\eta}_n \cdot V^{1/q},$$

and, by relation (4.18), $\hat{\eta}_n \cdot V^{1/q}$ is almost surely bounded. Therefore $\frac{m_n}{n} \cdot V^{1/q}$ is almost surely uniformly bounded. □

4.3.5 Examples of application of Theorem 4.33

We present two examples of applications. One of the more challenging aspects of applying the theorem is verifying assumption (c), which concerns the existence of such a quasi-stationary distribution (QSD) ν. In this regard, we rely on Foster-Lyapunov criteria, as discussed in [MT93]. These criteria are necessary in establishing different types of stability for continuous-time Markov processes, and provide a systematic framework for determining the existence of a QSD. While we will not provide a detailed insight of these criteria, we will apply them in the following examples. The study of existence and uniqueness of a QSD is relatively new and involves advanced mathematical tools. For further reading on this topic, one might refer to [CV23].

Example 4.40 ($M/M/\infty$ queue)**.** Consider a Markov chain defined on $E = \mathbb{N} = \{0, 1, 2, \dots\}$ with two positive parameters λ and μ such that $0 < \lambda < \mu$. The transition probabilities of the chain are given by

$$p_{x,x+1} = \frac{\lambda}{x\mu + \lambda} \quad \text{and} \quad p_{x,x-1} = \frac{x\mu}{x\mu + \lambda} \qquad (\forall x \geq 1),$$

with $p_{0,1} = 1$. This Markov chain is known as the $M/M/\infty$ queue, and it is known to be ergodic, and with stationary distribution

$$\pi(x) = \left(\frac{\lambda}{\mu}\right)^x \frac{e^{-\frac{\lambda}{\mu}}}{x!},$$

for every $x \geq 0$ (see e.g. section 2.5 of [S+16], or section 1.6 in [MM17]). In the context of measure-valued processes, the transition kernel R_x of the MVPP m_n on $\mathbb{N}$ is expressed as

$$R_x = \frac{\lambda}{x\mu + \lambda}\delta_{x+1} + \frac{x\mu}{x\mu + \lambda}\delta_{x-1},$$

for all $x \geq 1$ and $R_0 = \delta_1$. Notice that this MVPP is balanced, as $R_x(E) = 1$ for every $x \in E$. By Theorem 4.9, we have that, as $n \to \infty$,

$$\frac{m_n}{n} \xrightarrow{\mathbb{P}} \pi.$$

We now want to show that this convergence is actually almost sure, applying Theorem 4.33. Assumption (a) holds since $R_x(E) = 1$, for every $x \in \mathbb{N}$, and we can take $\rho = \delta_1$, so that $c = 1$. For assumption (b), notice that, the $R^{(i)}$'s are deterministic, hence $R^{(i)} = R$. Observe that, taking $V(x) = e^x$, the set $\{x \in \mathbb{N} : e^x \leq L\}$ is relatively compact for every $L \geq 0$: indeed, as a subset of $\mathbb{N}$, which is Hausdorff, it is enough to show that it is compact, i.e. bounded and closed. For $L = 0$, $\{x \in \mathbb{N} : e^x \leq 0\} = \varnothing$, which is compact, and for $L > 0$, $\{x \in \mathbb{N} : e^x \leq L\} = \{0, 1, \dots, \lfloor \log(L) \rfloor\}$, it is straightforward to see that it is bounded and closed. Hence, assumption (b-i) holds. We also have,

$$R_x \cdot V = \frac{\lambda}{x\mu + \lambda}e^{x+1} + \frac{x\mu}{x\mu + \lambda}e^{x-1} = \frac{\lambda e^2 + x\mu}{e(x\mu + \lambda)}V(x).$$

Notice that, it holds

$$\frac{\lambda e^2 + x\mu}{e(x\mu + \lambda)} < \frac{2}{e} \iff x > \frac{\lambda(e^2-2)}{\mu},$$

so $R_x \cdot V \leq \theta V(x) + K$, where $\theta = \frac{2}{e} \in (0, c) = (0, 1)$, and $K = \sup_{x \leq \lambda(e^2-2)/\mu} R_x \cdot V$, which means that (b-ii) is satisfied.
For part (iii), for every $r, p > 1$, we have

$$\mathbb{E}\big[R_x^{(1)}(E)^r\big] \vee \mathbb{E}\big[R_x^{(1)}(E)^p\big] = R_x(E)^r \vee R_x(E)^p = 1.$$

Since the $M/M/\infty$ queue is ergodic, we deduce that the continuous-time Markov process of generator $R - \mathrm{Id}$ is ergodic and the domain of attraction of π is $\mathcal{P}(\mathbb{N})$. Arguing as in the proof of Lemma 4.34, we get that, for any $q > 1$, $R_x \cdot V^{1/q} \leq \theta^{1/q} V(x) + K^{1/q}$, where $\theta^{1/q} < 1$. This and the Foster-Lyapunov type criteria of [MT93] provide the uniform convergence to π, which satisfies assumption (c). Assumption (d) follows by the fact that $E = \mathbb{N}$ is discrete. Therefore, by Theorem 4.33, if $\sum_{k \geq 0} e^k m_0(k) < \infty$, we get

$$\frac{m_n}{n} \xrightarrow{n \to \infty} \pi \quad \text{a.s.}$$

Example 4.41 (Out-degree of a RRT)**.** Recall the notion of random recursive tree discussed in section 4.2.2 (see e.g. Definition 4.13). We now define the *out-degree profile* of a rooted tree T as

$$\mathrm{Out}(T) = \sum_{\xi \in T} \delta_{\mathrm{outdeg}(\xi)},$$

where for all ξ in T, $\mathrm{outdeg}(\xi)$ is the out-degree of ξ, i.e. its number of children.
It is straightforward to see that the sequence $(\mathrm{Out}(\mathrm{RRT}_n))_{n \geq 0}$ is an MVPP of initial composition $m_0 = \delta_0$ and (deterministic) replacement kernel

$$R_x = \delta_{x+1} - \delta_x + \delta_0 \quad (\forall x \geq 0).$$

Indeed, at time $n = 0$, RRT_0 is composed only by the root (node 0), which has an out-degree of 0; when the tree evolves, at each time step $n + 1$, a child is added to a node ξ chosen uniformly at random from RRT_n. Supposing that at time n, $\mathrm{outdeg}(\xi) = x$, at time $n + 1$ the out-degree of ξ will increase by 1, and hence the count of nodes with an out-degree of x will decrease by 1. Moreover, the count of nodes with out-degree 0 will increase by 1, as a new node is added to the tree.
Notice that the replacement measures are not positive, so the process falls into the framework of assumptions (a), (b'), (c), (d), (e). Indeed, assumption (e) is satisfied by definition. As in the previous example, assumption (a) is satisfied as $R_x(\mathbb{N}) = 1$ for every natural x, and we can take $\rho = \delta_1$, so that $c = 1$.
Fix $\varepsilon \in (0, \frac{1}{2})$ and let $V(x) = (2-\varepsilon)^x$, for all $x \geq 0$. Arguing as in the previous example, the set

$$\big\{x \in \mathbb{N} : (2-\varepsilon)^x \leq L\big\} = \left\{0, 1, \ldots, \left\lfloor \frac{\log(L)}{\log(2-\varepsilon)} \right\rfloor \right\}$$

is relatively compact for every $L \geq 0$, so assumption (b'-i) holds. For part (ii), we have

$$R_x \cdot V = (2-\varepsilon)^{x+1} - (2-\varepsilon)^x + 1 = 1 + (2-\varepsilon)^x(1-\varepsilon) = 1 + (1-\varepsilon)V(x),$$

and for every $q \in (1,2]$,

$$R_x \cdot V^{1/q} = 1 + \big((2-\varepsilon)^{1/q} - 1\big)V(x)^{1/q} \leq 1 + (1-\varepsilon)V^{1/q},$$

since $\frac{1}{q} < 1$ and $2 - \varepsilon > 1$. Thus, assumption (b'-ii) holds for $\theta = 1 - \varepsilon$ and $K = 1$. Now notice that, for every function $f : \mathbb{N} \to \mathbb{R}$ bounded by 1, and for all $q' \in (1,3]$,

$$|R_x \cdot f|^{q'} = |f(x+1) - f(x) + f(0)|^{q'} \leq 3^{q'} \leq 27V(x),$$

since $V(x) \geq 1$ for all $x \geq 0$, which means that also part (iii) holds, with $A = 27$. Also,

$$|R_x \cdot V| = 1 + (1-\varepsilon)V(x) \leq 1 + (2-\varepsilon)V(x),$$

since $V(x) \geq 1$, and for every $q \in (1,2]$,

$$\left|R_x \cdot V^{1/q}\right|^q \leq \big(1 + (1-\varepsilon)V^{1/q}\big)^q \leq 2^q V(x),$$

thus (b'-iv) is satisfied. Notice that p is arbitrary in $(2,\infty)$, so q is arbitrary in $(1,2)$ and $q' \in (q,3]$.
One can verify that the Markov chain of kernel $(R_x)_{x\in\mathbb{N}}$ is ergodic, with unique stationary distribution $\upsilon(x) = 2^{-x-1}$, for every $x \geq 0$. As in the previous example, by [MT93], we have the uniform convergence to υ, which satisfies assumption (c), and, as before, (d) follows since $E = \mathbb{N}$ is discrete. Therefore, we can apply Theorem 4.33, to get, if $\sum_{k\geq 0}(2-\varepsilon)^k m_0(k) < \infty$, it holds

$$\frac{\mathrm{Out}(\mathrm{RRT}_n)}{n} \xrightarrow{n\to\infty} \upsilon \quad \text{a.s.}$$

Chapter 5

Conclusion

This thesis has explored the behaviour and convergence properties of Pólya urn models, extending classical results to broader scenarios, including Measure-Valued Pólya Processes (MVPPs). Through a combination of martingale theory and Markov chains, it provides a detailed insight of both finite and infinite colour cases. Alongside these results, it introduces the reader to further methods such as the embedding argument for urn models into continuous time of Athreya and Karlin ([AK68]), random recursive trees, branching Markov processes, stochastic approximations, and quasi-stationary distributions, forming a comprehensive introduction to various methods in probability theory.

In the case of finitely many colours, the study rigorously analysed the behaviour of the urn process under both identity and irreducible replacement matrices, providing insights into stochastic and deterministic convergence. These results were complemented by discussions on rates of convergence and fluctuations, deepening the understanding of the dynamics of reinforcement models. For the MVPP case, the generalization to infinite colour spaces opens up new possibilities, though there are still challenges, especially in proving results with fewer assumptions. For example, Theorem 4.33 shows that current results often rely on very specific conditions.

There are many possible directions for future research. For finitely many colours, the study of urns with random replacement matrices or extending results to non-diagonalizable matrices represents a natural progression. Most of these generalizations are already covered by Janson in [Jan04], who achieved results with fewer assumptions, but additional exploration in this area could further refine and generalize the theory. For MVPPs, this relatively young area remains rich with open questions. In particular, proving almost-sure convergence for non-balanced MVPPs with fewer assumptions would represent a significant theoretical advance; for instance, Theorem 4.33 illustrates how heavily current results rely on restrictive conditions.

Ultimately, this thesis aims to provide a structured and accessible entry point into these diverse topics. The combination of discrete and continuous models, random structures, and long-term behaviour provides an exciting area for further research.

Bibliography

[AGS23] Pantelis P Analytis, Alexandros Gelastopoulos, and Hrvoje Stojic. Ranking-based rich-get-richer processes. *The Annals of Applied Probability*, 33(6A):4366–4394, 2023.

[AIJA85] David J Aldous, Illdar A Ibragimov, Jean Jacod, and David J Aldous. *Exchangeability and related topics*. Springer, 1985.

[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.

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

[Ath69a] Krishna Balasundaram Athreya. Limit theorems for multitype continuous time markov branching processes: I. the case of an eigenvector linear functional. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, 12:320–332, 1969.

[Ath69b] Krishna Balasundaram Athreya. Limit theorems for multitype continuous time markov branching processes: II. the case of an arbitrary linear functional. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, 13(3):204–214, 1969.

[BCV21] Michel Benaïm, Nicolas Champagnat, and Denis Villemonais. Stochastic approximation of quasi-stationary distributions for diffusion processes in a bounded domain. *Annales de l'Institut Henri Poincaré, Probabilités et Statistiques*, 57(2):726 – 739, 2021.

[Ben06] Michel Benaïm. Dynamics of stochastic approximation algorithms. In *Seminaire de probabilites XXXIII*, pages 1–68. Springer, 2006.

[Ber40] S Bernstein. Nouvelles applications des grandeurs aléatoires presqu' indépendantes. *Izvestiya Rossiiskoi Akademii Nauk. Seriya Matematicheskaya*, 4(2):137–150, 1940.

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

[Bor21] Konstantin Borovkov. Gaussian process approximations for multicolor Pólya urn models. *Journal of Applied Probability*, 58(1):274–286, 2021.

[BP85] Amitava Bagchi and Asim K Pal. Asymptotic normality in the generalized Pólya–Eggenberger urn model, with an application to computer data structures. *SIAM Journal on Algebraic Discrete Methods*, 6(3):394–405, 1985.

[BST04] Michel Benaïm, Sebastian J. Schreiber, and Pierre Tarrès. Generalized Urn models of evolutionary processes. *The Annals of Applied Probability*, 14(3):1455 – 1478, 2004.

[Cha69] Srishti D Chatterji. An L^p-convergence theorem. *The Annals of Mathematical Statistics*, 40(3):1068–1070, 1969.

[CMP13] Brigitte Chauvin, Cécile Mailler, and Nicolas Pouyanne. Smoothing equations for large Pólya urns, 2013.

[CMSM13] Pierre Collet, Servet Martínez, and Jaime San Martín. Quasi-stationary distributions. *Markov chains, diffusions and dynamical systems*, 2013.

[CN14] Alla Dita Raza Choudary and Constantin P Niculescu. *Real analysis on intervals*. Springer, 2014.

[CPS11] Brigitte Chauvin, Nicolas Pouyanne, and Reda Sahnoun. Limit distributions for large Pólya urns. *The Annals of Applied Probability*, 21(1), February 2011.

[CV23] Nicolas Champagnat and Denis Villemonais. General criteria for the study of quasi-stationarity. *Electronic Journal of Probability*, 28(none):1 – 84, 2023.

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

[Duf13] Marie Duflo. *Random iterative models*, volume 34. Springer Science & Business Media, 2013.

[Dur19] Rick Durrett. *Probability: theory and examples*, volume 49. Cambridge university press, 2019.

[EP23] Florian Eggenberger and George Pólya. Über die statistik verkettteter vorgänge. *ZAMM-Journal of Applied Mathematics and Mechanics/Zeitschrift für Angewandte Mathematik und Mechanik*, 3(4):279–289, 1923.

[Fer73] Thomas S Ferguson. A Bayesian analysis of some nonparametric problems. *The annals of statistics*, pages 209–230, 1973.

[FGW+12] S Favaro, Alessandra Guglielmi, S Walker, et al. A class of measure-valued Markov chains and Bayesian nonparametrics. *BERNOULLI*, 18(3):1002–1030, 2012.

[FPS21] Sandra Fortini, Sonia Petrone, and Hristo Sariev. Predictive constructions based on measure-valued Pólya urn processes. *Mathematics*, 9(22):2845, 2021.

[Fre65] David A Freedman. Bernard friedman's urn. *The Annals of Mathematical Statistics*, pages 956–970, 1965.

[Fri49] Bernard Friedman. A simple urn model. *Communications on Pure and Applied Mathematics*, 2(1):59–70, 1949.

[Gra94] Ronald L Graham. *Concrete mathematics: a foundation for computer science*. Pearson Education India, 1994.

[GS20] Geoffrey Grimmett and David Stirzaker. *Probability and random processes*. Oxford university press, 2020.

[Gut13] A. Gut. *Probability: A Graduate Course*. Springer Texts in Statistics. Springer New York, 2013.

[HAG17] Mikhail Hayhoe, Fady Alajaji, and Bahman Gharesifard. A Pólya urn-based model for epidemics on networks. In *2017 American Control Conference (ACC)*, pages 358–363. IEEE, 2017.

[Han06] Chris Hand. History matters: Modelling path dependence on a spreadsheet. *Computers in Higher Education Economics Review*, 18(1):19–24, 2006.

[Har02] T.E. Harris. *The Theory of Branching Processes*. Dover phoenix editions. Dover Publications, 2002.

[HJ85] Roger A. Horn and Charles R. Johnson. *Matrix Analysis*. Cambridge University Press, 1985.

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

[JK77] Norman L. Johnson and Samuel Kotz. *Urn models and their applications: an approach to modern discrete probability theory*. John Wiley & Sons, 1977.

[Kal97] Olav Kallenberg. *Foundations of modern probability*, volume 2. Springer, 1997.

[KBJ04] Samuel Kotz, Narayanaswamy Balakrishnan, and Norman L Johnson. *Continuous Multivariate Distributions, Volume 1: Models and Applications*, volume 1. John Wiley & Sons, 2004.

[KK96] David G Kleinbaum and Mitchel Klein. *Survival analysis a self-learning text*. Springer, 1996.

[Lév37] P Lévy. Théorie de l'addition des variables aléatoires. *Gauthier-Villars, Paris*, 1937.

[LG22] Jean-François Le Gall. *Measure theory, probability, and stochastic processes*, volume 295. Springer Nature, 2022.

[Lin02] Torgny Lindvall. *Lectures on the coupling method*. Courier Corporation, 2002.

[LK20] Pong Lung Lau and Tay TR Koo. Online popularity of destinations in Australia: An application of Pólya urn process to search engine data. *Journal of Hospitality and Tourism Management*, 42:277–285, 2020.

[LM95] Douglas Lind and Brian Marcus. *An Introduction to Symbolic Dynamics and Coding*. Cambridge University Press, 1995.

[Mah91] Hosam M Mahmoud. Limiting distributions for path lengths in recursive trees. *Probability in the Engineering and Informational Sciences*, 5(1):53–59, 1991.

[Mah08] H. Mahmoud. *Pólya Urn Models*. Chapman and Hall/CRC, 2008.

[Mai23] Cécile Mailler. Pólya urns and other reinforcement processes. 2023.

[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.

[MM17] Cécile Mailler and Jean-François Marckert. Measure-valued Pólya urn processes. *Electronic Journal of Probability*, 22(none):1 – 33, 2017.

[MT93] Sean P Meyn and Richard L Tweedie. Stability of Markovian processes III: Foster–Lyapunov criteria for continuous-time processes. *Advances in Applied Probability*, 25(3):518–548, 1993.

[MV12] Sylvie Méléard and Denis Villemonais. Quasi-stationary distributions and population processes. *Probability Surveys*, 9:340 – 410, 2012.

[MV20] Cécile Mailler and Denis Villemonais. Stochastic approximation on noncompact measure spaces and application to measure-valued Pólya processes. *The Annals of Applied Probability*, 30(5):2393 – 2438, 2020.

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

[Pit06] Jim Pitman. *Combinatorial stochastic processes*. Springer, 2006.

[Ros14] Sheldon M Ross. *Introduction to probability models*. Academic press, 2014.

[Row31] Charles H. Rowe. A proof of the asymptotic series for logΓ(z) and logΓ(z+a). *Annals of Mathematics*, 32(1):10–16, 1931.

[RY13] Daniel Revuz and Marc Yor. *Continuous martingales and Brownian motion*, volume 293. Springer Science & Business Media, 2013.

[S+16] János Sztrik et al. *Basic queueing theory*. GlobeEdit OmniScriptum GmbH, KG, Saarbrucken, Germany, 2016.

[Sri08] Sashi Mohan Srivastava. *A course on Borel sets*, volume 180. Springer Science & Business Media, 2008.

[Ste01] J Michael Steele. *Stochastic calculus and financial applications*, volume 1. Springer, 2001.

[Szy90] Jerzy Szymanski. On the maximum degree and the height of a random recursive tree. In *Random graphs*, volume 87, pages 313–324, 1990.