ALMOST SURE AND MOMENT CONVERGENCE FOR TRIANGULAR PÓLYA URNS

SVANTE JANSON

Abstract. We consider triangular Pólya urns and show under very weak conditions a general strong limit theorem of the form $X_{ni}/a_{ni} \xrightarrow{\text{a.s.}} \mathcal{X}_i$, where X_{ni} is the number of balls of colour i after n draws; the constants a_{ni} are explicit and of the form $n^\alpha \log^\gamma n$; the limit is a.s. positive, and may be either deterministic or random, but is in general unknown.

The result extends to urns with subtractions under weak conditions, but a counterexample shows that some conditions are needed.

For balanced urns we also prove moment convergence in the main results if the replacements have the corresponding moments.

The proofs are based on studying the corresponding continuous-time urn using martingale methods, and showing corresponding results there. In the main part of the paper, we assume for convenience that all replacements have finite second moments; in an appendix this is relaxed to L^p for some $p > 1$.

Contents

Date: 21 March, 2024.
2020 *Mathematics Subject Classification.* 60F15, 60F25; 60C05, 60G44, 60J85.
Supported by the Knut and Alice Wallenberg Foundation and the Swedish Research Council (VR)
.

1. Introduction

A (generalized) Pólya urn contains balls of different colours. A ball is drawn at random from the urn, and is replaced by a set of balls that depends on the colour of the drawn balls; more generally, the replacement set may be random, with a distribution depending on the drawn colour. We assume that the set Q of colours is finite, and let $q := |\mathsf{Q}|$ be the number of colours. (See e.g. [31], [38], [27] for the history and further references.) It is often convenient to assume that $\mathsf{Q} = \{1, \dots, q\}$, but for us it will be convenient not to assume this.

The Pólya urn process can be defined formally as follows. The composition of the urn at time n is given by the vector $\mathbf{X}_n = (X_{ni})_{i \in \mathsf{Q}} \in [0, \infty)^q$, where X_{ni} is the number of balls of colour i. The urn starts with a given vector $\mathbf{X}_0$, and evolves according to a discrete-time Markov process. Each colour i has an *activity* $a_i \geqslant 0$, and a (generally random) *replacement vector* $\boldsymbol{\xi}_i = (\xi_{ij})_{j \in \mathsf{Q}}$. At each time $n + 1 \geqslant 1$, the urn is updated by drawing one ball at random from the urn, with the probability of any ball proportional to its activity. (In many cases $a_i = 1$ for all i, so all balls are drawn with equal probability; the reader may concentrate on this case until Section 5.) Thus, the drawn ball has colour i with probability

$$\frac{a_i X_{ni}}{\sum_j a_j X_{nj}}. \tag{1.1}$$

If the drawn ball has colour i, it is replaced together with $\xi_{ij}^{(n)}$ balls of colour j, $j \in \mathsf{Q}$, where the random vector $\boldsymbol{\xi}_i^{(n)} = (\xi_{ij}^{(n)})_{j \in \mathsf{Q}}$ is a copy of $\boldsymbol{\xi}_i$ that is independent of everything else that has happened so far. Thus, the urn is updated to

$$\mathbf{X}_{n+1} = \mathbf{X}_n + \boldsymbol{\xi}_i^{(n)}. \tag{1.2}$$

Remark 1.1. Note that, as in many other papers on Pólya urns, we do not assume that X_{ni} are integers; any real numbers $X_{ni} \geqslant 0$ are allowed. In general, it is thus a misnomer to call X_{ni} the "number" of balls of colour i; it is more precise to regard X_{ni} as the amount of colour i in the urn. Nevertheless, we will continue to use the traditional terminology, which thus has to be interpreted liberally by the reader. △

Remark 1.2. Since the drawn ball is replaced, it is also really a misnomer to call $\boldsymbol{\xi}_i$ the "replacement" vector; it is really an *addition* vector. (The replacements are really $\xi_{ij} + \delta_{ij}$.) Nevertheless, we use the terminology above, which is used in many papers. △

Remark 1.3. We allow the replacement vectors $\boldsymbol{\xi}_i$ to be random. (Some papers consider only the special case of deterministic $\boldsymbol{\xi}_i$, which is an important special case that appears in many applications.) We may say that the urn has *deterministic (or non-random) replacements* if all ξ_{ij} are deterministic, and otherwise *random replacements*. These terms should thus be interpreted as conditioned on (the colour of) the drawn ball. △

Remark 1.4. It is often convenient to describe the replacements by the *replacement matrix* $(\xi_{ij})_{i,j\in\mathsf{Q}}$. Note, however, that unless the replacements are deterministic, this may be somewhat misleading, since the rows should be regarded as separate random vectors, not necessarily defined on the same probability space; there is no need for a joint distribution of different rows. △

Remark 1.5. In the first part of the paper we assume that the replacements $\xi_{ij} \geqslant 0$ (Condition (A5) below), meaning that we only add balls to the urn and never remove any. In Section 8, and also usually in later sections, we more generally allow also that balls may be removed from the urn (assuming some hypotheses). △

Remark 1.6. We allow some activities a_i to be 0; this means that balls of colour i never are drawn. See Section 5 for an important example of this. (If all $a_i > 0$, we may reduce to the standard case $a_i = 1$ by considering the urn $(a_i X_{ni})_i$, with corresponding replacement vectors $(a_j \xi_{ij})_j$, but we will not use this.) △

Remark 1.7. We assumed tacitly above that the denominator $\sum_j a_j X_{nj}$ in (1.1) is > 0 for every $n \geqslant 0$, so that the definition makes sense. This holds, for example, under assumptions (A1) and (A5) below. (Urns that do not satisfy this, and therefore may stop at some finite time, have also been studied, but they will not be considered here.) △

We are interested in asymptotic properties of $\mathbf{X}_n$ as $n \to \infty$.

In the present paper, we study *triangular urns*, i.e., Pólya urns such that, for a suitable labelling of the colours by $1, \dots, q$, we have $\xi_{ij} = 0$ when $i > j$. (See also Section 2.3.) This includes the original Pólya urns studied by Markov [39], Eggenberger and Pólya [16], and Pólya [41] (all for $q = 2$), where the replacement matrix is *diagonal*: $\xi_{ij} = 0$ when $i \neq j$, but also many other interesting cases. See Section 14 for some examples.

There are many previous papers on triangular urns; we mention here only a few that are particularly relevant to the present paper; see also the references in the examples in Section 14. Athreya [2] studied diagonal urns with random replacements and showed a.s. convergence of the proportions of different colours, using the embedding mthod of Athreya and Karlin [3] that is also the basis of the present paper. Gouet [18, 19] proved (in particular) an a.s. convergence result for triangular urns with 2 colours and deterministic replacements, assuming also that the urn is *balanced*, meaning $\sum_j \xi_{ij} = b$ for some constant b (and all $a_i = 1$, see further Section 10). Janson [28] studied triangular urns with 2 colours and deterministic $\boldsymbol{\xi}_i$, and proved convergence in distribution (but not a.s.) of the components X_{ni} after suitable normalizations; there are several cases, and the limits are sometimes normal and sometimes not. This was partially extended by Aguech [1], who also studied triangular urns with 2 colours, but allowed random replacements $\boldsymbol{\xi}_i$ (under some hypotheses, see Example 14.4); moreover, he proved convergence a.s., and not just in distribution. Bose, Dasgupta, and Maulik [12] (and [11] for $q = 2$) studied triangular urns with an arbitrary (finite) number of colours; they assumed that the replacements are deterministic, and that the urn is balanced, and then, under some further assumptions, showed convergence a.s. of the components X_{ni}, suitably normalized, see Example 14.12.

The main purpose of the present paper is to extend these results by Gouet [18, 19], Aguech [1], and Bose, Dasgupta, and Maulik [12], and show a.s. convergence for triangular urns with any (finite) number of colours, allowing replacements $\boldsymbol{\xi}_i$ that are both random and unbalanced. Our main result is the following, using the technical assumptions (A1)–(A5) in Section 2.1 and the notation defined in (2.7)–(2.13) in Section 2.4 below. See also the extensions Theorems 8.4 and B.1 where the technical assumptions are weakened (allowing urns with subtractions and reducing our moment assumption, respectively); since the proof of the theorem is rather long and we want to focus on the main ideas, we use first the conditions (A1)–(A5) (which suffice for many applications), and add later the extra arguments needed for the extensions.

Theorem 1.8. *Let $(X_{ni})_{i\in\mathsf{Q}}$ be a triangular Pólya urn satisfying the conditions* (A1)–(A5) *below. Then, for every colour $i \in \mathsf{Q}$, there exists a random variable $\widehat{\mathcal{X}}_i$ with $0 < \widehat{\mathcal{X}}_i < \infty$ a.s. such that as $n \to \infty$:*

(i) *If $\widehat{\lambda} > 0$, then*

$$\frac{X_{ni}}{n^{\lambda_i^*/\widehat{\lambda}} \log^{\gamma_i} n} \xrightarrow{\text{a.s.}} \widehat{\mathcal{X}}_i. \tag{1.3}$$

(ii) *If $\widehat{\lambda} = 0$, then*

$$\frac{X_{ni}}{n^{\kappa_i/\widehat{\kappa}_0}} \xrightarrow{\text{a.s.}} \widehat{\mathcal{X}}_i. \tag{1.4}$$

Note that the exponent γ_i may be both positive and negative; see the examples in Section 14. Note also that the various exponents in (1.3) and (1.4) are explicitly given in Section 2.4, but the limiting random variables $\widehat{\mathcal{X}}_i$ are known only in some special cases; in general they are unfortunately unknown.

The virtue of Theorem 1.8 is that it is very general, but as discussed in the remaks below, more precise results are known in some special cases.

In Theorem 1.8, the main case is (i), $\widehat{\lambda} > 0$, and the reader should focus on this case. The case $\widehat{\lambda} = 0$ is more special and of less interest for applications, but it is included for completeness; by (2.10) and (2.7), this case occurs when $\xi_{ii} = 0$ for every $i \in \mathsf{Q}$. (We might call such urns *strictly triangular*.)

Another major result in the present paper (Theorem 12.5) says in particular that in the special case of balanced triangular urns, if the replacements have finite moments, then the a.s. limits in Theorem 1.8 hold also in L^p for any $p < \infty$, and thus moments converge. (It is an open problem whether this extends to some class of unbalanced triangular urns.)

Remark 1.9. Theorem 1.8 describes the first-order asymptotics of the urn. We will see in Section 7 that the limiting random variable $\widehat{\mathcal{X}}_i$ is deterministic (i.e., a constant) in some cases, but not in general. In cases where Theorem 1.8 yields a limit $\widehat{\mathcal{X}}_i$ that is deterministic (and perhaps also otherwise), it is interesting to study fluctuations (i.e., second order terms) and try to find limits (e.g. in distribution, after a suitable normalization) for the difference of the two sides of (1.3) or (1.4). Such results in some cases are given in [19], [28] and [1]; see the examples in Section 14, but we will not pursue this problem here, and leave it as an open problem. △

Remark 1.10. Convergence almost surely implies convergence in distribution. Thus, as a corollary, (1.3) holds also with convergence in distribution. However, our proof does not seem to provide a method to find the limit distribution, i.e. the distribution of $\widehat{\mathcal{X}}_i$, except in some very simple cases. Moreover, the limits $\widehat{\mathcal{X}}_i$ are (in general) dependent.

For $q = 2$ and deterministic ξ_{ij}, limit distributions were given in [28]. (Sometimes degenerate, sometimes not.) The results there thus describe the distribution of $\widehat{\mathcal{X}}_i$ in this case, although the descriptions in some cases are complicated. Some further examples of known limit distributions are given in some examples in Section 14. We leave the general case as another open problem. △

The proof of Theorem 1.8 is given in Sections 3–5 below, after some preliminaries in Section 2. The proof is based on the embedding by Athreya and Karlin [3] of a Pólya urn into a continuous-time multitype branching process (Section 2.6); we then apply martingale methods to obtain a continuous-time version of Theorem 1.8 (Theorem 4.1); finally, this implies results for the embedded discrete-time urn. The proof is generalized to urns with subtraction in Section 8, and to urns with a weaker moment condition in Appendix B. Since the proofs are rather long and technical, we prefer to first present the proof in the basic case Theorem 1.8 (which is enough for most applications) and later discuss the modifications required for the extensions, instead of proving the most general results immediately.

Remark 1.11. Gouet [18, 19] and Bose, Dasgupta, and Maulik [12], in special cases (see Example 14.3 and Example 14.12), instead study X_{ni} directly and use martingale metods in discrete time. It seems that this approach (also used by several authors for non-triangular urns) works well for balanced urns, but that the embedding into continuous time works better for unbalanced urns. △

Remark 1.12. We consider in this paper only triangular Pólya urns. Another important class of urns consists of the irreducible urns. In this case a.s. convergence (under some technical conditions) was shown by Athreya and Karlin [3], see also [4, Section V.9.3] and [27, Theorem 3.21].

It might be possible to combine the methods of the present paper and the methods for irreducible urns to obtain results on a.s. convergence for all types of Pólya urns (under some technical conditions), see Remark 2.3, but the present paper is long as it is and we leave this as a speculation for future research. (Note also the counterexamples Example 14.14 and 14.15, showing that some conditions are needed even in the triangular case.) △

1.1. **Contents.** Section 2 contains preliminaries, including some definitions and notation. Section 3 consists of a series of lemmas that comprise the main technical part of our proofs. They lead to the main theorem for continuous time in Section 4, which in turn is used to prove Theorem 1.8 in Section 5. Sections 6 (continuous time) and 7 (discrete time) contain results on whether the limit random variable are degenerate (i.e., constant) or not, and some related results.

Section 8 extend the previous results to urns with subtracion, where we allow $\xi_{ii} = -1$. With some extra technical conditions, the previous results hold in this case too, with only minor modifications of the proofs.

The following sections contain some complements. Section 9 is a short section comparing random and non-random replacements with the same means.

Section 10 contains some general results on balanced urn, mainly as preliminaries to the following sections.

Section 11 considers the number of times a given colour is drawn; it is shown that the results of earlier sections extend to this case.

Section 12 contains results on convergence in L^2 and in L^p, and closely related results on convergence of moments in, for example, Theorem 1.8. For the main result (discrete-time), we have to assume that urn is balanced, and we state an open problem for more general urns.

Section 13 discusses briefly another open problem (rates of convergence).

Section 14 contains a number of examples that illustrate the results and their limitations, and also give connections to previous literature.

Finally, Appendix A contains some simple general lemmas on absolute continuity that we believe are known, but for which we were unable to find references. Appendix B gives proofs of L^p versions of L^2 estimates used in the main part of the paper; this yields both the extension of Theorem 1.8 mentioned above, and a proof of the results in Section 12. Appendix C gives a rather technical proof of one claim in Example 14.14.

2. Some notation and other preliminaries

We use throughout the paper the notation Q, q, $\mathbf{X}_n = (X_{ni})_{i\in\mathsf{Q}}$, a_i, and $\boldsymbol{\xi}_i = (\xi_{ij})_{j\in\mathsf{Q}}$ introduced in the introduction.

2.1. **Standing assumptions.** In the rest of the paper we assume

(A0) The Pólya urn is triangular.

(Unless we explicitly say so, for example when we discuss this property in Section 2.3.) For the central part of the paper (Sections 2–7) we make also some standing technical assumptions:

(A1) The initial urn $\mathbf{X}_0$ is non-random. Moreover, each $X_{0i} \geqslant 0$ and $\sum a_i X_{0i} > 0$. (The results may be extended to random $\mathbf{X}_0$ by conditioning on $\mathbf{X}_0$.)

(A2) If $a_i = 0$, then $\boldsymbol{\xi}_i = 0$, i.e., $\xi_{ij} = 0$ for every $j \in \mathsf{Q}$. (This is without loss of generality, since $a_i = 0$ means that balls of colour i never are drawn, and thus $\boldsymbol{\xi}_i$ does not matter.)

(A3) For every $i \in \mathsf{Q}$, either $X_{0i} > 0$, or there exists $j \neq i$ such that $\mathbb{P}(\xi_{ji} > 0) > 0$ (or both). (This too is without loss of generality, since otherwise balls of colour i can never appear, so $X_{ni} = 0$ a.s. for all n, and we may remove the colour i from Q.)

(A4) $\mathbb{E}\,\xi_{ij}^2 < \infty$ for all $i, j \in \mathsf{Q}$.

(A5) $\xi_{ij} \geqslant 0$ (a.s.) for all $i, j \in \mathsf{Q}$.

Note that (A5) implies that every X_{ni} is (weakly) increasing in n. In particular, the urn never gets empty. Combined with (A1) we see that $\sum_i a_i X_{ni} > 0$ for every n, and thus the probabilities (1.1) and the urn process are well defined. We discuss extensions to urns not satisfying (A5) (i.e., urns with subtractions) in Section 8; see Theorems 8.4–8.6.

Remark 2.1. The second moment condition (A4) is for technical convenience. In fact, the results (including Theorem 1.8) hold assuming only $\mathbb{E}\,\xi_{ij}^p < \infty$ for some $p > 1$. However, this adds further arguments to an already long proof, so we assume second moments in the main part of the paper and show the extension to $p > 1$ in Appendix B.

It seems possible that the results could extend further by assuming only that $\mathbb{E}\,\xi_{ij} \log \xi_{ij} < \infty$, as for diagonal urns in [2] using related results for (single-type) branching processes [4, Theorem III.7.2]. (The results do not extend without modifications to cases without this assumption, see Example 14.16.) We have not pursued this, and leave it as an open problem. △

2.2. **General notation.** As usual, we ignore events of probability 0. We often write "almost surely" or "a.s." for emphasis, but we may also tacitly omit this.

We use "increasing" in the weak sense. Similarly, "positive function" means in the weak sense, i.e. $\geqslant 0$. (However, "positive constant" always means strictly positive; we sometimes add "strictly" for emphasis, but not always.)

We let $s \wedge t := \min\{s, t\}$ and $s \vee t := \max\{s, t\}$.

We let $\mathbb{Z}_{\geqslant 0} := \{0, 1, 2, \dots\}$ and $\mathbb{Z}_+ := \{1, 2, \dots\}$.

We use $\xrightarrow{\text{a.s.}}$, $\xrightarrow{\text{p}}$, and $\xrightarrow{\text{d}}$ to denote convergence almost surely, in probability, and in distribution, respectively.

"Absolutely continuous" (for a probability distribution in $\mathbb{R}$ or $\mathbb{R}^d$) means with respect to Lebesgue measure. (Except in Appendix A when a reference measure is explicitly specified.) We may say that a random variable is absolutely continuous when its distribution is.

We use some standard probability distributions: $\mathrm{Exp}(\lambda)$ is an exponential distribution with mean $\lambda > 0$; we may also say with *rate* $1/\lambda$. $\Gamma(\alpha, b)$ is a Gamma distribution. (Thus $\mathrm{Exp}(\lambda) = \Gamma(1, \lambda)$.) $\mathrm{Be}(p)$ is a Bernoulli distribution. $\mathrm{NegBin}(r, p)$ is a negative binomial distribution. $\mathrm{Ge}(p)$ is a geometric distribution on $\{1, 2, \dots\}$.

For a random variable W, we let $\mathcal{L}(W)$ denotes the distribution of W, and, for any $p > 0$,

$$\|W\|_p := \bigl(\mathbb{E}\,|W|^p\bigr)^{1/p}. \tag{2.1}$$

L^p denotes the set of all random variables W such that $\|W\|_p < \infty$.

C denotes unspecified constants that may vary from one occurrence to the next. They may depend on the activities a_i and the distributions of the replacements $\boldsymbol{\xi}_i$, and perhaps on other parameters clear from the context, but they never depend on n or t.

Let

$$r_{ij} := \mathbb{E}\,\xi_{ij}. \tag{2.2}$$

Thus the matrix $(r_{ij})_{i,j\in\mathsf{Q}}$ is the mean replacement matrix. Note that r_{ij} exists and is finite by (A4). By (A5), we have $r_{ij} \geqslant 0$, and $r_{ij} = 0 \iff \xi_{ij} = 0$ a.s.

We occasionally (when we discuss two different urns at the same time) denote a Pólya urn by $\mathcal{U}$; we then (somewhat informally) mean both the urn process $\mathbf{X}_n$ and its continuous-time version $\mathbf{X}(t)$ defined below, and also the colour set Q and the replacement matrix (ξ_{ij}).

2.3. **The colour graph.** (In this subsection, the urn does not have to be triangular.) Recall that Q is the set of colours. As said in the introduction, we allow Q to be any finite set, although it is possible to assume $\mathsf{Q} = \{1, \dots, q\}$ without loss of generality when this is convenient.

We regard the set Q of colours as a directed graph, called the *colour graph*, where for any distinct $i, j \in \mathsf{Q}$ there is an edge $i \to j$ if and only if $\mathbb{P}(\xi_{ij} \neq 0) > 0$. In other words, $i \to j$ means that if a ball of colour i is drawn, it is possible (with positive probability) that some balls of colour j are added. Note that by (A2), $i \to j$ entails $a_i > 0$, so balls of colour i may (and will) actually be drawn; furthermore, by (A5),

$$i \to j \iff r_{ij} > 0 \text{ and } i \neq j. \tag{2.3}$$

We say, again for two distinct colours $i, j \in \mathsf{Q}$, that i is an *ancestor* of j, and j a *descendant* of i, if there exists a directed path in Q from i to j; we denote this by $i \prec j$. In other words, $i \prec j$ means that if we start the urn with only a ball of colour i, then it is possible that balls of colour j are added at some later time.

If $i \in \mathsf{Q}$, let $\mathsf{P}_i := \{j \in \mathsf{Q} : j \to i\}$, the set of colours different from i whose drawings may cause addition of balls of colour i. We say that the colour i is *minimal* if $\mathsf{P}_i = \varnothing$. We denote the set of minimal colours by $\mathsf{Q}_{\min}$. Note that (A3) can be formulated as: $X_{0i} > 0$ for every $i \in \mathsf{Q}_{\min}$.

We say that the urn is *triangular*, if there exists a (re)labelling of the colours by $1, \dots, q$ that makes the matrix $(\xi_{ij})_{i,j\in\mathsf{Q}}$ triangular a.s. (Assuming (A5), this is equivalent to the mean replacement matrix $(r_{ij})_{i,j\in\mathsf{Q}}$ being triangular, cf. (2.3).) In other words, the urn is triangular if there exists a total ordering $<$ of the colours such that

$$i > j \implies \xi_{ij} = 0 \quad \text{a.s.} \tag{2.4}$$

Using the definitions above to rewrite (2.4), we see that the urn is triangular if and only there exists a total ordering $<$ such that, for $i, j \in \mathsf{Q}$,

$$i \to j \implies i < j. \tag{2.5}$$

Furthermore, this is equivalent to

$$i \prec j \implies i < j. \tag{2.6}$$

Proposition 2.2. *The following are equivalent.*

(i) *The urn is triangular.*
(ii) *The colour graph is acyclic.*
(iii) *The relation $\prec$ on Q is a partial order.*

Proof. (i) implies (ii) and (iii) as a consequence of (2.5) and (2.6).

(ii) $\iff$ (iii) is easily seen.

Finally, any partial order can be extended to a total order. Thus, if (iii) holds, we may extend $\prec$ to a total order $<$, which means that (2.6) holds. ∎

Note that if the urn is triangular, a colour i is minimal if and only if it is minimal in the partial order $\prec$.

Remark 2.3. We may also note that a Pólya urn is irreducible if and only if its colour graph is strongly connected. Given any Pólya urn, we may decompose its colour graph into its strongly connected components, which are linked by the remaining edges in an acyclic way. Hence the urn can be regarded as an acyclic directed network of irreducible urns. This suggests, as mentioned in Remark 1.12, that the methods in the present paper perhaps might be combined with methods for irreducible urns to obtain results for general urns. △

2.4. **More notation.** Let, for $i \in \mathsf{Q}$,

$$\lambda_i := a_i r_{ii} = a_i \mathbb{E}\,\xi_{ii} \geqslant 0. \tag{2.7}$$

In the continuous-time version introduced below, λ_i is the rate of additions to colour i by drawings of the same colour. Since colours also may be added by drawings of another colour, we define further

$$\lambda_i^* := \max\bigl\{\lambda_j : j \preceq i\bigr\} \geqslant 0. \tag{2.8}$$

In other words, λ_i^* is the largest λ_j for a colour j such that there exists a path (possibly of length 0) in Q from j to i. Such a path may contain several colours k with the same, maximal, λ_k, and we denote the largest number of them in a single path by $1 + \kappa_i$; i.e.,

$$\kappa_i := \max\bigl\{\kappa : \exists i_1 \prec i_2 \prec \cdots \prec i_{\kappa+1} \preceq i \text{ with } \lambda_{i_1} = \cdots = \lambda_{i_{\kappa+1}} = \lambda_i^*\bigr\} \geqslant 0. \tag{2.9}$$

Define further

$$\widehat{\lambda} := \max\{\lambda_i : i \in \mathsf{Q}\} \geqslant 0, \tag{2.10}$$

$$\begin{aligned}\widehat{\kappa} &:= \max\{\kappa_i : i \in \mathsf{Q} \text{ and } \lambda_i^* = \widehat{\lambda}\}\\ &= \max\bigl\{\kappa : \exists i_1 \prec i_2 \prec \cdots \prec i_{\kappa+1} \text{ with } \lambda_{i_1} = \cdots = \lambda_{i_{\kappa+1}} = \widehat{\lambda}\bigr\} \geqslant 0.\end{aligned} \tag{2.11}$$

If $\widehat{\lambda} = 0$ (i.e., if $\lambda_i = 0$ for every $i \in \mathsf{Q}$), let further

$$\widehat{\kappa}_0 := 1 + \max\bigl\{\kappa_i : i \in \mathsf{Q} \text{ with } a_i > 0\bigr\} \geqslant 1. \tag{2.12}$$

If $\widehat{\lambda} > 0$, define also

$$\gamma_i := \kappa_i - \widehat{\kappa}\lambda_i^*/\widehat{\lambda}, \qquad i \in \mathsf{Q}. \tag{2.13}$$

2.5. **Stochastic processes.** All our continuous-time stochastic processes are defined on $[0, \infty)$ and are assumed to be càdlàg (right-continuous with left limits).

We consider martingales without explicitly specifying the filtration; this will always be the natural filtration $(\mathcal{F}_t)_{t\geqslant 0}$, where $\mathcal{F}_t$ is generated by "everything that has happened up to time t". If T is a stopping time, then $\mathcal{F}_T$ denotes the corresponding σ-field generated by all events up to time T.

Given a stochastic process $W = (W(t))_{t\geqslant 0}$, we define its maximal process by

$$W^*(t) := \sup_{0\leqslant s\leqslant t} |W(s)|, \qquad 0 \leqslant t \leqslant \infty. \tag{2.14}$$

(We consider only $s < \infty$, also when $t = \infty$.) We further define

$$\Delta W(t) := W(t) - W(t-), \qquad 0 \leqslant t < \infty, \tag{2.15}$$

where $W(t-)$ is the left limit at t, with $W(0-) := 0$.

If W is a process with locally bounded variation, we may define its *quadratic variation* by

$$[W,W]_t := \sum_{0\leqslant s\leqslant t} \bigl|\Delta W(s)\bigr|^2, \qquad 0 \leqslant t \leqslant \infty, \tag{2.16}$$

(summing over $s < \infty$ if $t = \infty$), noting that the sum always really is countable. (We have no need for the definition for general semimartingales, see e.g. [32, p. 519 and Theorem 26.6] or [42, Section II.6].) Recall [42, Corollary 3 to Theorem II.6.27, p. 73] that if M is a local martingale and $\mathbb{E}\,[M,M]_t < \infty$ for some $t < \infty$, then

$$\mathbb{E}\,|M(t)|^2 = \mathbb{E}\,[M,M]_t, \tag{2.17}$$

and as a consequence, if $\mathbb{E}\,[M,M]_\infty < \infty$, then M is an L^2-bounded martingale and thus $M(\infty) := \lim_{t\to\infty} M(t)$ exists a.s., and (2.17) holds for all $t \leqslant \infty$.

Recall also Doob's inequality [32, Proposition 7.16] which as a special case yields, combined with (2.17),

$$\mathbb{E}\,M^*(t)^2 \leqslant C\,\mathbb{E}\,|M(t)|^2 = C\,\mathbb{E}\,[M,M]_t, \tag{2.18}$$

provided $\mathbb{E}\,[M,M]_t < \infty$. (Here $C = 4$, but we will not use this.)

2.6. **Continuous-time urn.** We will use the standard method of embedding the discrete-time urn in a continuous-time process, due to Athreya and Karlin [3], see also [4, §V.9], and after them used in many papers. We thus define the *continuous-time urn* as a vector valued Markov process $\mathbf{X}(t) = ((X_i(t))_{i\in\mathsf{Q}}$ with given initial value $\mathbf{X}(0) := \mathbf{X}_0$ such that, for each $i \in \mathsf{Q}$, "a ball of colour i is drawn" with intensity $a_iX_i(t)$; when a ball of colour i is drawn, we add to $\mathbf{X}(t)$ a copy of $\boldsymbol{\xi}_i$ (independent of the history). In the classical case (e.g. [3]) when each $X_i(t)$ is integer valued, $\mathbf{X}(t)$ is a multitype continuous-time Markov branching process; in general (allowing any real $X_i(t) \geqslant 0$), $\mathbf{X}(t)$ is a (vector-valued) continuous-time continuous-state branching process (abbreviated *CB process*) as defined by Jiřina [30], see also e.g. Li [37]. (The process $\mathbf{X}(t)$ is of jump-type, as in [30, Section 3].) Note that (A4) implies $\mathbb{E}\,\xi_{ij} < \infty$ for all $i, j \in \mathsf{Q}$, which is a well-known sufficient condition for non-explosions; i.e., there exists such a process $\mathbf{X}(t)$ with $\mathbf{X}(t)$ finite for all $t \in [0,\infty)$.

Since $a_iX_i(0) > 0$ for some i by (A1), and $X_i(t)$ is increasing by (A5), there will a.s. be infinitely many draws in the urn. We let $\widehat{T}_n$ be the nth time that a ball is drawn, with $\widehat{T}_0 := 0$. Then the discrete-time urn $\mathbf{X}_n$ in Section 1 can be realized as

$$\mathbf{X}_n := \mathbf{X}(\widehat{T}_n). \tag{2.19}$$

We assume (2.19) throughout the paper.

Since $\mathbf{X}(t)$ does not explode, there is a.s. only a finite number of draws up to any finite time t, and thus $\widehat{T}_n \xrightarrow{\text{a.s.}} \infty$ as $n \to \infty$. Note that we denote the discrete-time urn by $\mathbf{X}_n = (X_{ni})_i$ and the continuous-time urn by $\mathbf{X}(t) = (X_i(t))_i$.

Of course, the continuous-time urn can also be studied for its own sake, see for example [13].

3. Analysis of one colour

In this section, we study one fixed colour $i \in \mathsf{Q}$ in the continuous-time urn.

Recall that $X_i(t)$ is the number (amount) of balls of colour i at time t in the continuous-time urn. There are several possible sources of these balls: some may be there from the beginning, some may be added when a ball of some other colour j is drawn (with $j \to i$ and thus $j \prec i$), and in both cases these balls of colour i may later be drawn and produce further generations of balls of colour i (provided $\lambda_i > 0$).

We begin by considering, in the following two subsections, two simpler special cases. We will then in the final subsection, rather easily, treat the general case by combining these special cases.

We first state some simple general results; these are more or less known, see e.g. [27, Lemma 9.3] for a related result, but for convenience we give full proofs.

Fix a colour j. Let $0 < T_1 < T_2 < \dots$ be the times that a ball of colour j is drawn, and let $N(t) := \bigl|\{k : T_k \leqslant t\}\bigr|$ denote the corresponding counting process; i.e., $N(t)$ is the number of draws of colour j up to time t. These draws occur with intensity $a_j X_j(t)$, which means that

$$\widetilde{N}(t) := N(t) - a_j \int_0^t X_j(s)\,\mathrm{d}s, \qquad t \geqslant 0, \tag{3.1}$$

is a local martingale. In fact, $\widetilde{N}(t)$ is a martingale, which we verify by the following simple lemma. Recall the notation (2.14).

Lemma 3.1. *Suppose that* $\mathbb{E}\, X_j(t) < \infty$ *for some* $t \in [0,\infty)$. *Then,*

$$\mathbb{E}\, N(t) < \infty, \tag{3.2}$$

$$\mathbb{E}\, \widetilde{N}^*(t) = \mathbb{E} \sup_{s \leqslant t} |\widetilde{N}(s)| < \infty. \tag{3.3}$$

In particular, if $\mathbb{E}\, X_j(t) < \infty$ *for every* $t < \infty$, *then the local martingale* $\widetilde{N}(t)$ *is a martingale.*

Proof. By the definition of local martingale, there exists an increasing sequence of stopping times τ_m, $m \geqslant 1$, such that $\tau_m \nearrow \infty$ a.s. as $m \to \infty$, and $\widetilde{N}(t \wedge \tau_m)$ is a martingale for each m. In particular, since $\widetilde{N}(0) = 0$,

$$\mathbb{E}\, N(t \wedge \tau_m) = \mathbb{E}\, \widetilde{N}(t \wedge \tau_m) + a_j\, \mathbb{E} \int_0^{t\wedge\tau_m} X_j(s)\,\mathrm{d}s = a_j\, \mathbb{E} \int_0^{t\wedge\tau_m} X_j(s)\,\mathrm{d}s. \tag{3.4}$$

Since $N(t)$ and $X_j(t)$ are increasing positive functions of t, we may use monotone convergence and let $m \to \infty$ to obtain

$$\mathbb{E}\, N(t) = a_j\, \mathbb{E} \int_0^t X_j(s)\,\mathrm{d}s \leqslant a_j\, \mathbb{E}\bigl[t X_j(t)\bigr] < \infty. \tag{3.5}$$

The monotonicity of $N(t)$ further implies

$$\widetilde{N}^*(t) \leqslant N(t) + a_j \int_0^t X_j(s)\,\mathrm{d}s \leqslant N(t) + a_j t X_j(t), \tag{3.6}$$

and thus (3.3) follows by (3.5).

The final statement follows since a local martingale with integrable maximal function is a martingale. □

Lemma 3.2. *Suppose that* $\mathbb{E}\, X_j(t) < \infty$ *for some* $t \in [0,\infty)$.

(i) *Let* f *be a positive or bounded measurable function on* $[0,t]$. *Then*

$$\mathbb{E} \sum_{k=1}^{\infty} \mathbf{1}_{\{T_k \leqslant t\}} f(T_k) = \mathbb{E} \int_0^t f(s)\,\mathrm{d}N(s) = a_j\, \mathbb{E} \int_0^t f(s) X_j(s)\,\mathrm{d}s. \tag{3.7}$$

(ii) *Let also* $(\eta_k)_1^\infty$ *be a sequence of identically distributed random variables with finite mean* $\mathbb{E}\,\eta_1$ *such that* η_k *is independent of* T_k. *Then*

$$\mathbb{E}\sum_{k=1}^{\infty}\mathbf{1}_{\{T_k\leqslant t\}}f(T_k)\eta_k = a_j\,\mathbb{E}\,\eta_1\,\mathbb{E}\int_0^t f(s)X_j(s)\,\mathrm{d}s. \tag{3.8}$$

Proof. Step 1. The left and middle terms in (3.7) are the same, since $\int_0^t f(s)\,\mathrm{d}N(s) = \sum_{k=1}^\infty \mathbf{1}_{\{T_k\leqslant t\}}f(T_k)$ by the definition of N.

Step 2. Suppose that $f(s) = \mathbf{1}_{\{a<s\leqslant b\}}$, the indicator function of an interval $(a,b] \in [0,t]$. Then, by (3.1),

$$\int_0^t f(s)\,\mathrm{d}N(s) - a_j\int_0^t f(s)X_j(s)\,\mathrm{d}s = \int_0^t f(s)\,\mathrm{d}\widetilde{N}(s) = \widetilde{N}(b) - \widetilde{N}(a), \tag{3.9}$$

and $\mathbb{E}\,[\widetilde{N}(b) - \widetilde{N}(a)] = 0$ by Lemma 3.1; hence (3.7) holds for such f.

Step 3. The monotone class theorem [21, Theorem 1.2.3] now shows that (3.7) holds for the indicator function $f(s) = \mathbf{1}_{\{s\in A\}}$ of any Borel set $A \in [0,t]$.

Step 4. By linearity, (3.7) holds for any positive simple function f. Then, by monotone convergence, (3.7) holds for any positive measurable function, and by linearity again for any bounded measurable function. This proves (i).

Step 5. In (ii), we may decompose η_k into its positive and negative parts; thus it suffices to consider $\eta_k \geqslant 0$. Then the sum in (3.8) is well-defined, and

$$\mathbb{E}\sum_{k=1}^{\infty}\mathbf{1}_{\{T_k\leqslant t\}}f(T_k)\eta_k = \sum_{k=1}^{\infty}\mathbb{E}\big[\mathbf{1}_{\{T_k\leqslant t\}}f(T_k)\eta_k\big] = \mathbb{E}\,\eta_1\sum_{k=1}^{\infty}\mathbb{E}\big[\mathbf{1}_{\{T_k\leqslant t\}}f(T_k)\big], \tag{3.10}$$

and (3.8) follows from (3.7). ∎

3.1. **A colour not influenced by others.** In this subsection, we assume that $\xi_{ji} = 0$ a.s. for all colours $j \neq i$. Equivalently, $j \not\to i$ for $j \in \mathsf{Q}$, i.e., $\mathsf{P}_i = \varnothing$; in other words, i is a minimal colour. This means that $X_i(t)$ is affected only by draws of the same colour i, and we may thus ignore all other colours and regard $X_i(t)$ as a continuous-time urn with a single colour. In other words, $X_i(t)$ is a one-dimensional CB process, starting at some given $X_i(0)$ and adding copies of ξ_{ii} with intensity $a_iX_i(t)$. We write $X_i(0) = x_0$. Note that our assumption (A3) means $x_0 > 0$, but for completeness we allow also the trivial case $x_0 = 0$ in the present subsection. (In this subsection, we really use only the assumptions (A1), (A4), and (A5), and only for the colour i.) Note also that $\lambda_i^* = \lambda_i$ by (2.8).

We will only need a few simple facts about CB processes; see further e.g. [30; 36; 20; 9; 37] where many more results are given.

Recall that $\lambda_i = a_ir_{ii}$. If $\lambda_i = 0$, then there are no additions at all, and $X_i(t) = X_i(0) = x_0$ is constant.

It is easy to see that since $r_{ii} < \infty$, the CB process $X_i(t)$ is well defined and non-explosive (i.e., finite for all t), for any given $x_0 \geqslant 0$. Moreover [20], [9, Proposition 2.2],

$$\mathbb{E}\,X_i(t) = e^{\lambda_i t}x_0. \tag{3.11}$$

This implies by conditioning and the Markov property that [4, Theorem III.7.1]

$$e^{-\lambda_i t}X_i(t) \quad \text{is a martingale.} \tag{3.12}$$

The following result too well known, but we include a proof for completeness, and since we discuss modifications of it later. See e.g. [4, Theorem V.8.2] and [27, Lemma 9.5] for a vector-valued extension.

Lemma 3.3. *Suppose that* $\xi_{ji} = 0$ *a.s. for all* $j \neq i$. *Then* $e^{-\lambda_i t} X_i(t)$ *is an* L^2-*bounded martingale, and thus*

$$e^{-\lambda_i t} X_i(t) \xrightarrow{\text{a.s.}} \mathcal{X}_i \tag{3.13}$$

for some random variable $\mathcal{X}_i$. *Furthermore, with* $x_0 = X_i(0)$,

$$\mathbb{E}\,\mathcal{X}_i = x_0, \tag{3.14}$$

$$\operatorname{Var}\mathcal{X}_i = x_0 \frac{a_i}{\lambda_i}\,\mathbb{E}\,\xi_{ii}^2, \tag{3.15}$$

where we interpret $\frac{0}{0}$ *as* 0. *Hence*

$$\operatorname{Var}\bigl[e^{-\lambda_i t} X_i(t)\bigr] \leqslant \operatorname{Var}\mathcal{X}_i \leqslant C x_0, \tag{3.16}$$

$$\mathbb{E}\,\Bigl|\sup_{t\geqslant 0} e^{-\lambda_i t} X_i(t)\Bigr|^2 \leqslant C\,\mathbb{E}\,\mathcal{X}_i^2 \leqslant C(x_0 + x_0^2). \tag{3.17}$$

Furthermore, if $x_0 > 0$, *then* $0 < \mathcal{X}_i < \infty$ *a.s.*

Proof. The case $\lambda_i = 0$ is trivial, with $\mathcal{X}_i = x_0$. The same holds if $x_0 = 0$. We may thus assume $\lambda_i > 0$ and $x_0 > 0$.

We argue as in [27, Section 9]. As above (now taking $j = i$), let $0 < T_1 < T_2 < \dots$ be the times that a ball of colour i is drawn, and let $N(t) := \bigl|\{i : T_i \leqslant t\}\bigr|$ denote the corresponding counting process- Furthermore, let $\eta_k := \Delta X_i(T_k)$ be the number of balls of colour i added at the k-th draw. Thus $\eta_1, \eta_2, \dots$ is a sequence of independent copies of ξ_{ii}.

Let $M(t) := e^{-\lambda_i t} X_i(t)$. Then M is a martingale by (3.12), and its quadratic variation is by (2.16) given by

$$[M,M]_t = \sum_{0\leqslant s\leqslant t} |\Delta M(s)|^2 = x_0^2 + \sum_{T_k\leqslant t} e^{-2\lambda_i T_k} |\Delta X_i(T_k)|^2 = x_0^2 + \sum_{T_k\leqslant t} e^{-2\lambda_i T_k}\eta_k^2. \tag{3.18}$$

Hence, it follows from Lemma 3.2(ii) that

$$\mathbb{E}\,[M,M]_t = x_0^2 + a_i\,\mathbb{E}\,[\xi_{ii}^2]\int_0^t e^{-2\lambda_i s}\,\mathbb{E}\,X_i(s)\,\mathrm{d}s. \tag{3.19}$$

Hence, writing $\beta := \mathbb{E}\,\xi_{ii}^2$, (3.11) yields

$$\mathbb{E}\,[M,M]_t = x_0^2 + a_i\beta x_0 \int_0^t e^{-\lambda_i s}\,\mathrm{d}s = x_0^2 + x_0\frac{a_i}{\lambda_i}\beta\bigl(1 - e^{-\lambda_i t}\bigr). \tag{3.20}$$

This shows by (2.17) that M is an L^2-bounded martingale; thus (3.13) holds for some $\mathcal{X}_i = M(\infty)$. Clearly $\mathcal{X}_i \geqslant 0$. We have $\mathbb{E}\,M(\infty) = \mathbb{E}\,M_0 = x_0$, which yields (3.14). Furthermore, (3.15) holds by (2.17) and (3.20) again, together with (3.14); this yields also (3.16). Doob's inequality (2.18) yields (3.17).

Finally, the distribution of $\mathcal{X}_i$ depends on x_0; thus let us denote $\mathcal{X}_i$ by $\mathcal{X}_i(x_0)$. The CB property implies that if $x, y \geqslant 0$, then $\mathcal{X}_i(x+y) \overset{\mathrm{d}}{=} \mathcal{X}_i(x) + \mathcal{X}_i'(y)$, where $\mathcal{X}_i'(y)$ is a copy of $\mathcal{X}_i(y)$ independent of $\mathcal{X}_i(x)$. Let $p(x) := \mathbb{P}\bigl(\mathcal{X}_i(x) = 0\bigr) \in [0,1]$. It follows that for any $x, y \geqslant 0$,

$$p(x+y) = p(x)p(y), \tag{3.21}$$

and thus there exists $c \in [0, \infty]$ such that

$$\mathbb{P}(\mathcal{X}_i(x) = 0) = p(x) = e^{-cx}, \qquad x > 0. \tag{3.22}$$

We must have $c > 0$, since otherwise $\mathcal{X}_i = 0$ a.s. for any x_0, which contradicts (3.14). The Markov property and (3.22) yield, for any $t \in [0, \infty)$,

$$\mathbb{P}\big(\mathcal{X}_i = 0 \mid X_i(t)\big) = e^{-cX_i(t)}. \tag{3.23}$$

Thus, taking the expectation,

$$\mathbb{P}\big(\mathcal{X}_i = 0\big) = \mathbb{E}\, e^{-cX_i(t)}. \tag{3.24}$$

It is clear that if $x_0 > 0$ and $\lambda_i > 0$, so $a_i > 0$ and $\mathbb{E}\,\xi_{ii} > 0$, then there is a.s. an infinite number of draws, and the (standard) law of large numbers shows that $X_i(t) \xrightarrow{\text{a.s.}} \infty$. Consequently, letting $t \to \infty$ in (3.23) yields, by dominated convergence,

$$\mathbb{P}(\mathcal{X}_i = 0) = 0. \tag{3.25}$$

□

By (3.15), the limit $\mathcal{X}_i$ is degenerate only in the trivial cases when $x_0 = 0$ or $a_i = 0$. We note that except in these cases, the limit has an absolutely continuous distribution.

Lemma 3.4. *In Lemma 3.3, suppose further that $\lambda_i > 0$ and $x_0 > 0$. Then the distribution of $\mathcal{X}_i$ is absolutely continuous.*

Proof. Let T_1 be the first time that a ball of colour i is drawn; then $T_1 \in \text{Exp}(1/(a_i x_0))$. The distribution of the balls added at T_1 is independent of T_1, and thus $X_i(T_1)$ is independent of T_1. It follows by the strong Markov property that the stochastic process $Y(t) := X_i(T_1 + t)$, $t \geqslant 0$, is independent of T_1. By (3.13), we have as $t \to \infty$,

$$e^{-\lambda_i t} Y(t) = e^{\lambda_i T_1} e^{-\lambda_i (T_1 + t)} X_i(T_1 + t) \xrightarrow{\text{a.s.}} \mathcal{Y} := e^{\lambda_i T_1} \mathcal{X}_i, \tag{3.26}$$

and it follows that $\mathcal{Y}$ is independent of T_1. Consequently,

$$\mathcal{X}_i = e^{-\lambda_i T_1} \mathcal{Y}, \tag{3.27}$$

where the two factors on the right-hand side are independent. The result follows from (3.27) since $e^{-\lambda_i T_1}$ has an absolutely continuous distribution and $\mathcal{Y} > 0$ a.s. (because $\mathcal{X}_i > 0$ a.s. by Lemma 3.3). □

3.2. **A colour only produced by one other colour.** We continue to consider a fixed colour i. In this subsection we assume that $|\mathsf{P}_i| = 1$; thus there is exactly one colour $j \in \mathsf{Q}$ such that $j \to i$. We assume also that $X_i(0) = 0$, so there are initially no balls of colour i. Recall that the assumption $j \to i$ means $r_{ij} > 0$, and thus implicitly $a_j > 0$.

Since the colour graph is acyclic, there is no feedback from colour i to colour j; thus we may regard the entire process $X_j(t)$, $t \in [0, \infty)$, as known, and consider only its effect on $X_i(t)$.

We may regard $X_i(t)$ as a CB process with immigration. Let again $0 < T_1 < T_2 < \dots$ be the times that a ball of colour j is drawn, and let $\eta_1, \eta_2, \dots$ be the corresponding number (amount) of balls of colour i that are added. In case there is only a finite number K of times that a ball of colour j is drawn, we define for completeness $T_k = \infty$ for $k > K$, and pick η_k with the correct distribution and independent of everything else. We regard each η_k as an immigrant coming at T_k. (Some η_k may be 0; this is no problem.) We can separate the descendants of each of these immigrants, and write

$$X_i(t) = \sum_{k:T_k \leqslant t} Y_k(t - T_k) = \sum_{k=1}^{\infty} \mathbf{1}_{\{T_k \leqslant t\}} Y_k(t - T_k) \tag{3.28}$$

where each $Y_k(t)$, conditioned on η_k, is a copy of the single-colour CB process in Section 3.1 with $Y_k(0) = \eta_k$. Furthermore, conditionally on $(\eta_k)_k$, the processes $Y_k(t)$ are independent. Recall that the additions $\eta_1, \eta_2, \dots$ are i.i.d. copies of ξ_{ji}; in particular, they have expectations $\mathbb{E}\,\eta_k = r_{ji}$. The times T_k are stopping times; moreover, the process Y_k is independent of T_k and the σ-field $\mathcal{F}_{T_k}$.

The following lemma contains much of the technical parts of our argument. It makes an assumption on the growth of $X_j(t)$ that later will be justified by an induction argument.

We define, for $m \in \mathsf{Q}$,

$$\widetilde{X}_m(t) := t^{-\kappa_m} e^{-\lambda_m^* t} X_m(t), \qquad 0 < t < \infty, \tag{3.29}$$

$$\widetilde{X}_m^{**} := \sup_{t \geqslant 0} \bigl\{ (t+1)^{-\kappa_m} e^{-\lambda_m^* t} X_m(t) \bigr\}. \tag{3.30}$$

We use powers of t in (3.29) and of $t + 1$ in (3.30) for technical convenience below (and we therefore use a notation with ** instead of *); the difference is not important since we are mainly interested in limits as $t \to \infty$.

Lemma 3.5. *Let $i \in \mathsf{Q}$. Suppose that there is exactly one colour $j \in \mathsf{Q}$ such that $j \to i$, and that $X_i(0) = 0$. Suppose further that*

$$\widetilde{X}_j(t) \xrightarrow{\text{a.s.}} \mathcal{X}_j \qquad \text{as } t \to \infty, \tag{3.31}$$

for some random variable $\mathcal{X}_j \geqslant 0$, and

$$\bigl\| \widetilde{X}_j^{**} \bigr\|_2 < \infty. \tag{3.32}$$

Then

$$\widetilde{X}_i(t) \xrightarrow{\text{a.s.}} \mathcal{X}_i \qquad \text{as } t \to \infty, \tag{3.33}$$

for some $\mathcal{X}_i \geqslant 0$ and

$$\bigl\| \widetilde{X}_i^{**} \bigr\|_2 < \infty. \tag{3.34}$$

Furthermore, a.s., if $\mathcal{X}_j > 0$, then $\mathcal{X}_i > 0$.

Moreover, if $\lambda_i \leqslant \lambda_j^*$, and thus $\lambda_i^* = \lambda_j^*$, then

$$\mathcal{X}_i = \begin{cases} \frac{a_j r_{ji}}{\lambda_j^* - \lambda_i} \mathcal{X}_j, & \lambda_i < \lambda_j^*, \\ \frac{a_j r_{ji}}{\kappa_i} \mathcal{X}_j, & \lambda_i = \lambda_j^*. \end{cases} \tag{3.35}$$

Proof. Using (3.28), and recalling $Y_k(0) = \eta_k$, we make the decomposition

$$e^{-\lambda_i t} X_i(t) = \sum_{k=1}^{\infty} \mathbf{1}_{\{T_k \leqslant t\}} e^{-\lambda_i T_k} \bigl(e^{-\lambda_i (t - T_k)} Y_k(t - T_k) - Y_k(0) \bigr) \tag{3.36}$$

$$+ \sum_{k=1}^{\infty} \mathbf{1}_{\{T_k \leqslant t\}} e^{-\lambda_i T_k} (\eta_k - r_{ji}) \tag{3.37}$$

$$+ r_{ji}\Big(\sum_{k=1}^{\infty} \mathbf{1}_{\{T_k \leqslant t\}} e^{-\lambda_i T_k} - \int_0^t e^{-\lambda_i s} a_j X_j(s)\,\mathrm{d}s\Big) \tag{3.38}$$

$$+ r_{ji} a_j \int_0^t e^{-\lambda_i s} X_j(s)\,\mathrm{d}s \tag{3.39}$$

$$=: Z_1(t) + Z_2(t) + r_{ji} Z_3(t) + a_j r_{ji} Z_4(t), \tag{3.40}$$

say. We consider the four processes $Z_\ell(t)$ in (3.40) separately (initially, at least). Moreover, recalling (2.8) and (2.9), we consider sometimes separately the three cases:

(i) $\lambda_i > \lambda_j^*$: then $\lambda_i^* = \lambda_i$ and $\kappa_i = 0$.
(ii) $\lambda_i = \lambda_j^*$: then $\lambda_i^* = \lambda_i$, and $\kappa_i = \kappa_j + 1$.
(iii) $\lambda_i < \lambda_j^*$: then $\lambda_i^* = \lambda_j^*$, and $\kappa_i = \kappa_j$.

We define, in analogy with (3.29)–(3.30) (but note the different exponents $\lambda_i - \lambda_i^*$), for $\ell = 1, 2, 3, 4$,

$$\widetilde{Z}_\ell(t) := t^{-\kappa_i} e^{(\lambda_i - \lambda_i^*)t} Z_\ell(t), \tag{3.41}$$

$$\widetilde{Z}_\ell^{**} := \sup_{t \geqslant 0} \big\{ (t+1)^{-\kappa_i} e^{(\lambda_i - \lambda_i^*)t} |Z_\ell(t)| \big\}. \tag{3.42}$$

Then (3.40) yields

$$\widetilde{X}_i(t) = \widetilde{Z}_1(t) + \widetilde{Z}_2(t) + r_{ji} \widetilde{Z}_3(t) + a_j r_{ji} \widetilde{Z}_4(t), \tag{3.43}$$

$$\widetilde{X}_i^{**} \leqslant \widetilde{Z}_1^{**} + \widetilde{Z}_2^{**} + C \widetilde{Z}_3^{**} + C \widetilde{Z}_4^{**}. \tag{3.44}$$

We will prove, for $\ell = 1, \dots, 4$ and some $\mathcal{Z}_\ell$,

$$\widetilde{Z}_\ell(t) \xrightarrow{\text{a.s.}} \mathcal{Z}_\ell \qquad \text{as } t \to \infty, \tag{3.45}$$

$$\| \widetilde{Z}_\ell^{**} \|_2 < \infty, \tag{3.46}$$

and then (3.33)–(3.34) follow from (3.43)–(3.44).

We treat $Z_1(t), \dots, Z_4(t)$ in (3.39) in reverse order, partly because $Z_4(t)$ will turn out to be the main term (sometimes at least).

Step 1: Z_4. By (3.39) and (3.30),

$$0 \leqslant Z_4(t) \leqslant \widetilde{X}_j^{**} \int_0^t (s+1)^{\kappa_j} e^{(\lambda_j^* - \lambda_i)s}\,\mathrm{d}s$$

$$\leqslant \begin{cases} C \widetilde{X}_j^{**} = C(t+1)^{\kappa_i} \widetilde{X}_j^{**}, & \lambda_i > \lambda_j^*, \\ (t+1)^{\kappa_j + 1} \widetilde{X}_j^{**} = (t+1)^{\kappa_i} \widetilde{X}_j^{**}, & \lambda_i = \lambda_j^*, \\ C(t+1)^{\kappa_j} e^{(\lambda_j^* - \lambda_i)t} \widetilde{X}_j^{**}, & \lambda_i < \lambda_j^*. \end{cases} \tag{3.47}$$

It follows that in all three cases (i)–(iii), $\widetilde{Z}_4^{**} \leqslant C \widetilde{X}_j^{**}$. In particular, (3.32) implies

$$\big\| \widetilde{Z}_4^{**} \big\|_2 < \infty. \tag{3.48}$$

For (3.45) we treat the three cases (i)–(iii) separately, in each case using (3.41) and (3.29)–(3.30). First, in case (i), we let $t \to \infty$ in (3.39) and (3.47) and conclude that a.s.

$$\widetilde{Z}_4(t) = Z_4(t) \to Z_4(\infty) := \int_0^\infty e^{-\lambda_i s} X_j(s)\,\mathrm{d}s \leqslant C \widetilde{X}_j^{**} < \infty. \tag{3.49}$$

Hence

$$\widetilde{Z}_4(t) \xrightarrow{\text{a.s.}} \mathcal{Z}_4 = Z_4(\infty). \tag{3.50}$$

Note that if $\mathcal{X}_j > 0$, then $X_j(t) > 0$ for large t, and thus $\mathcal{Z}_4 = Z_4(\infty) > 0$ by (3.49).

In case (ii), we have for $t \geqslant 1$, using the change of variables $s = xt$,

$$\begin{aligned}\widetilde{Z}_4(t) &= t^{-\kappa_i} Z_4(t) = t^{-\kappa_j - 1} \int_0^t s^{\kappa_j} \widetilde{X}_j(s)\,\mathrm{d}s \\ &= t^{-\kappa_j-1} \int_0^1 s^{\kappa_j} \widetilde{X}_j(s)\,\mathrm{d}s + \int_{1/t}^1 x^{\kappa_j} \widetilde{X}_j(xt)\,\mathrm{d}x \\ &\to 0 + \int_0^1 x^{\kappa_j} \mathcal{X}_j\,\mathrm{d}x = (\kappa_j+1)^{-1}\mathcal{X}_j = \kappa_i^{-1}\mathcal{X}_j, \end{aligned} \tag{3.51}$$

by (3.31) and dominated convergence, which applies since (3.29)–(3.30) show that $\sup_{s\geqslant 1} \widetilde{X}_j(s) \leqslant 2^{\kappa_j} \widetilde{X}_j^{**} < \infty$ a.s.

In case (iii), similarly, now with $s = t - u$,

$$\begin{aligned}\widetilde{Z}_4(t) &= t^{-\kappa_j} e^{(\lambda_i - \lambda_j^*)t} \int_0^t s^{\kappa_j} e^{(\lambda_j^* - \lambda_i)s} \widetilde{X}_j(s)\,\mathrm{d}s \\ &= \int_0^t \Bigl(\frac{t-u}{t}\Bigr)^{\kappa_j} e^{-(\lambda_j^*-\lambda_i)u} \widetilde{X}_j(t-u)\,\mathrm{d}u \\ &\to \int_0^\infty e^{-(\lambda_j^*-\lambda_i)u} \mathcal{X}_j\,\mathrm{d}u = (\lambda_j^* - \lambda_i)^{-1}\mathcal{X}_j, \end{aligned} \tag{3.52}$$

using dominated convergence again, justified by, for $t \geqslant 1$,

$$\Bigl(\frac{t-u}{t}\Bigr)^{\kappa_j} \widetilde{X}_j(t-u) \leqslant \Bigl(\frac{t-u+1}{t}\Bigr)^{\kappa_j} \widetilde{X}_j^{**} \leqslant 2^{\kappa_j} \widetilde{X}_j^{**} < \infty. \tag{3.53}$$

We have thus shown (3.45) and (3.46) for $\ell = 4$ in all cases, with $\mathcal{Z}_4 > 0$ when $\mathcal{X}_j > 0$. Furthermore, in cases (ii) and (iii), we have

$$\mathcal{Z}_4 = \begin{cases} \kappa_i^{-1}\mathcal{X}_j, & \lambda_i = \lambda_j^*, \\ (\lambda_j^* - \lambda_i)^{-1}\mathcal{X}_j, & \lambda_i < \lambda_j^*. \end{cases} \tag{3.54}$$

Step 2: Z_3, first part. We have, recalling (3.1),

$$Z_3(t) = \int_0^t e^{-\lambda_i s}\,\mathrm{d}\widetilde{N}(s), \tag{3.55}$$

which is a local martingale since $\widetilde{N}(t)$ is; in fact it is a martingale since

$$Z_3^*(t) \leqslant N(t) + a_j \int_0^t e^{-\lambda_i s} X_j(s)\,\mathrm{d}s \leqslant N(t) + tX_j(t), \tag{3.56}$$

and thus $\mathbb{E}\, Z_3^*(t) < \infty$ by Lemma 3.1, noting that the assumption (3.32) implies $\mathbb{E}\, X_j(t) < \infty$ for all $t \geqslant 0$. Since $Z_3(t)$ has locally bounded variation and jumps only at the times T_k, with $\Delta Z_3(T_k) = e^{-\lambda_i T_k}$, its quadratic variation is

$$[Z_3, Z_3]_t = \sum_{k=1}^\infty \mathbf{1}_{\{T_k \leqslant t\}} e^{-2\lambda_i T_k} = \int_0^t e^{-2\lambda_i s}\,\mathrm{d}N(s). \tag{3.57}$$

The draws T_k occur with rate $a_j X_j(t)$, and it follows by (3.57) and Lemma 3.2 that, recalling (3.30),

$$\mathbb{E}\,|Z_3(t)|^2 = \mathbb{E}\,[Z_3, Z_3]_t = a_j\,\mathbb{E} \int_0^t e^{-2\lambda_i s} X_j(s)\,\mathrm{d}s$$

$$\leqslant a_j \, \mathbb{E}\, \widetilde{X}_j^{**} \int_0^t (s+1)^{\kappa_j} e^{(\lambda_j^* - 2\lambda_i)s} \, \mathrm{d}s. \tag{3.58}$$

The rest of the argument will be the same for Z_2 and Z_1, so we first consider them.

Step 3: Z_2, first part. Each term in the sum (3.37) is a martingale, since T_k is a stopping time, and $\eta_k - r_{ji}$ has mean 0 and is independent of the σ-field $\mathcal{F}_{T_k}$. Hence, the sum $Z_2(t)$ is at least a local martingale (since stopping at any T_k yields a finite sum and thus a martingale). Since $Z_2(t)$ jumps only at the times T_k, and is constant in between, its quadratic variation is given by

$$[Z_2, Z_2]_t = \sum_{k=1}^{\infty} \mathbf{1}_{\{T_k \leqslant t\}} e^{-2\lambda_i T_k} (\eta_k - r_{ji})^2. \tag{3.59}$$

Hence, again using the independence between η_k and T_k,

$$\begin{aligned} \mathbb{E}\, [Z_2, Z_2]_t &= \sum_{k=1}^{\infty} \mathbb{E}\big[\mathbf{1}_{\{T_k \leqslant t\}} e^{-2\lambda_i T_k}\big] \, \mathbb{E}\big[(\eta_k - r_{ji})^2\big] \\ &= \operatorname{Var}[\eta_1] \, \mathbb{E} \sum_{k=1}^{\infty} \mathbf{1}_{\{T_k \leqslant t\}} e^{-2\lambda_i T_k}. \end{aligned} \tag{3.60}$$

We recognize the final sum from (3.57), and conclude

$$\mathbb{E}\, [Z_2, Z_2]_t = C \, \mathbb{E}\, [Z_3, Z_3]_t. \tag{3.61}$$

Step 4: Z_1, first part. We write the sum in (3.36) as $Z_1(t) = \sum_{k=1}^{\infty} Z_1^{(k)}(t)$, with

$$Z_1^{(k)}(t) := \mathbf{1}_{\{t \geqslant T_k\}} e^{-\lambda_i T_k} \big(e^{-\lambda_i (t - T_k)} Y_k(t - T_k) - Y_k(0)\big). \tag{3.62}$$

It is easily seen that each $Z_1^{(k)}(t)$ is a martingale, since T_k is a stopping time and $e^{-\lambda_i t} Y_k(t) - Y_k(0)$ is a martingale starting at 0, which furthermore is independent of $\mathcal{F}_{T_k}$. Hence, for every finite $m \geqslant 1$, the finite sum

$$Z_1^{[\leqslant m]}(t) := \sum_{k=1}^{m} Z_1^{(k)}(t) \tag{3.63}$$

is a martingale, and thus $Z_1(t \wedge T_m) = Z_1^{[\leqslant m]}(t \wedge T_m)$ is a martingale. Consequently, Z_1 is a local martingale. Furthermore, conditioned on all T_k and η_k, the processes $Z_1^{(k)}(t)$ are independent and thus a.s. they jump at different times. (Note that $Z_1^{(k)}(0) = 0$.) Hence, by (2.16),

$$[Z_1, Z_1]_t = \sum_{k=1}^{\infty} [Z_1^{(k)}, Z_1^{(k)}]_t \tag{3.64}$$

and thus, using (3.62) and (2.17) again together with (3.16),

$$\begin{aligned} \mathbb{E}\big[[Z_1, Z_1]_t \mid (T_k, \eta_k)_1^{\infty}\big] &= \sum_{k=1}^{\infty} \mathbb{E}\big[[Z_1^{(k)}, Z_1^{(k)}]_t \mid (T_k, \eta_k)_1^{\infty}\big] \\ &= \sum_{k=1}^{\infty} \mathbf{1}_{\{T_k \leqslant t\}} e^{-2\lambda_i T_k} \operatorname{Var}\big[e^{-\lambda_i (t - T_k)} Y_k(t - T_k) \mid T_k, \eta_k\big] \\ &\leqslant \sum_{k=1}^{\infty} \mathbf{1}_{\{T_k \leqslant t\}} e^{-2\lambda_i T_k} C \eta_k. \end{aligned} \tag{3.65}$$

This yields, by taking the expectation,

$$\mathbb{E}\,[Z_1, Z_1]_t \leqslant C\,\mathbb{E}\sum_{k=1}^{\infty} \mathbf{1}_{\{T_k \leqslant t\}} e^{-2\lambda_i T_k}. \tag{3.66}$$

This is the same sum as in (3.57), and we conclude

$$\mathbb{E}\,[Z_1, Z_1]_t \leqslant C\,\mathbb{E}\,[Z_3, Z_3]_t. \tag{3.67}$$

Step 5: Z_1, Z_2, Z_3, final part. Let $\ell \in \{1,2,3\}$. In all three cases, $Z_\ell(t)$ is a local martingale such that, by (3.61), (3.67), and (3.58), for $t \in [0,\infty)$,

$$\mathbb{E}\,[Z_\ell, Z_\ell]_t \leqslant C\,\mathbb{E}\,[Z_3, Z_3]_t \leqslant C\,\mathbb{E}\,\widetilde{X}_j^{**} \int_0^t (s+1)^{\kappa_j} e^{(\lambda_j^* - 2\lambda_i)s}\,\mathrm{d}s. \tag{3.68}$$

In particular, (3.68) shows that $\mathbb{E}\,[Z_\ell, Z_\ell]_t < \infty$ for every $t < \infty$, and thus $Z_\ell(t)$ is a square integrable martingale, and $\mathbb{E}\,|Z_\ell(t)|^2 = \mathbb{E}\,[Z_\ell, Z_\ell]_t$.

We now consider another partition into three separate cases for λ_j^* and λ_i.

(i′) $\lambda_j^* < 2\lambda_i$. Then we can let $t \to \infty$ in (3.68) and obtain $\mathbb{E}\,[Z_\ell, Z_\ell]_\infty < \infty$. Consequently, $Z_\ell(t)$ is an L^2-bounded martingale, and thus

$$Z_\ell(t) \xrightarrow{\text{a.s.}} Z_\ell(\infty) < \infty, \qquad \text{as } t \to \infty, \tag{3.69}$$

and $Z_\ell^*(\infty) := \sup_{t \geqslant 0} |Z_\ell(t)| \in L^2$. Furthermore (as always), $\lambda_i^* \geqslant \lambda_i$, and thus (3.42) implies $\widetilde{Z}_\ell^{**} \leqslant Z_\ell^*(\infty)$; consequently, $\|\widetilde{Z}_\ell^{**}\|_2 < \infty$.

We consider two subcases, recalling the cases discussed at the beginning of the proof:

(i′a) *Case* (i), $\lambda_i > \lambda_j^*$. Then $\lambda_i^* = \lambda_i$ and $\kappa_i = 0$; hence (3.41) yields $\widetilde{Z}_\ell(t) = Z_\ell(t)$ and thus (3.69) yields $\widetilde{Z}_\ell(t) \to \mathcal{Z}_\ell := Z_\ell(\infty)$ a.s.

(i′b) *Case* (ii) *or* (iii), $\lambda_i \leqslant \lambda_j^*$. Then $\lambda_i^* > \lambda_i$ or $\kappa_i \geqslant 1$ and thus (3.41) and (3.69) yield $\widetilde{Z}_\ell(t) \to \mathcal{Z}_\ell := 0$ a.s.

(ii′) $\lambda_j^* \geqslant 2\lambda_i$ *and* $\lambda_j^* > 0$. Let, for $n \geqslant 1$,

$$\widetilde{Z}_\ell^\dagger(n) := \sup_{n-1 \leqslant t \leqslant n} (t+1)^{-\kappa_i} e^{(\lambda_i - \lambda_i^*)t} |Z_\ell(t)|. \tag{3.70}$$

It follows from Doob's inequality (2.18) and (3.68) that, with $\overline{C} := C\,\mathbb{E}\,\widetilde{X}_j^{**}$,

$$\begin{aligned}
\mathbb{E}\,\widetilde{Z}_\ell^\dagger(n)^2 &\leqslant C n^{-2\kappa_i} e^{2(\lambda_i - \lambda_i^*)n}\,\mathbb{E}\,Z_\ell^*(n)^2 \\
&\leqslant C n^{-2\kappa_i} e^{2(\lambda_i - \lambda_i^*)n}\,\mathbb{E}\,[Z_\ell, Z_\ell]_n \\
&\leqslant \overline{C} n^{-2\kappa_i} e^{2(\lambda_i - \lambda_i^*)n} \int_0^n (s+1)^{\kappa_j} e^{(\lambda_j^* - 2\lambda_i)s}\,\mathrm{d}s \\
&\leqslant \overline{C} n^{-2\kappa_i} e^{2(\lambda_i - \lambda_i^*)n} n^{\kappa_j + 1} e^{(\lambda_j^* - 2\lambda_i)n} \\
&= \overline{C} n^{\kappa_j + 1 - 2\kappa_i} e^{(\lambda_j^* - 2\lambda_i^*)n}.
\end{aligned} \tag{3.71}$$

We have always $\lambda_i^* \geqslant \lambda_j^*$, and thus, in the present case, $\lambda_j^* - 2\lambda_i^* \leqslant -\lambda_j^* < 0$. Hence (3.71) yields

$$\mathbb{E}\sum_{n=1}^{\infty} \widetilde{Z}_\ell^\dagger(n)^2 = \sum_{n=1}^{\infty} \mathbb{E}\,\widetilde{Z}_\ell^\dagger(n)^2 < \infty. \tag{3.72}$$

Consequently, a.s. $\widetilde{Z}_\ell^\dagger(n) \to 0$ as $n \to \infty$, which by (3.41) and (3.70) means $\widetilde{Z}_\ell(t) \to 0$ as $t \to \infty$. Moreover, (3.42) implies

$$\big(\widetilde{Z}_\ell^{**}\big)^2 \leqslant \sum_{n=1}^{\infty} \widetilde{Z}_\ell^\dagger(n)^2, \tag{3.73}$$

and thus (3.72) implies also $\|\widetilde{Z}_\ell^{**}\|_2 < \infty$.

(iii′) $\lambda_j^* = \lambda_i = 0$. In this case we have $\lambda_i^* = 0$ and $\kappa_i = \kappa_j + 1$. Let, for $n \geqslant 1$,

$$\widetilde{Z}_\ell^\ddagger(n) := \sup_{2^{n-1}\leqslant t\leqslant 2^n} t^{-\kappa_i} e^{(\lambda_i-\lambda_i^*)t}|Z_\ell(t)| = \sup_{2^{n-1}\leqslant t\leqslant 2^n} t^{-\kappa_j-1}|Z_\ell(t)|. \tag{3.74}$$

Similarly to (3.71), it follows from Doob's inequality and (3.68) that

$$\mathbb{E}\,\widetilde{Z}_\ell^\ddagger(n)^2 \leqslant \overline{C}2^{-2(\kappa_j+1)n}\int_0^{2^n}(s+1)^{\kappa_j}\,\mathrm{d}s \leqslant \overline{C}2^{-(\kappa_j+1)n}. \tag{3.75}$$

Hence,

$$\mathbb{E}\sum_{n=1}^{\infty}\widetilde{Z}_\ell^\ddagger(n)^2 = \sum_{n=1}^{\infty}\mathbb{E}\,\widetilde{Z}_\ell^\ddagger(n)^2 < \infty, \tag{3.76}$$

and the rest of the argument is as in the preceding case, now using

$$\big(\widetilde{Z}_\ell^{**}\big)^2 \leqslant Z_\ell^*(1)^2 + \sum_{n=1}^{\infty}\widetilde{Z}_\ell^\ddagger(n)^2 \tag{3.77}$$

and noting that $\mathbb{E}\,Z_\ell^*(1)^2 < \infty$ by (2.18) and (3.68).

We have shown that (3.45) and (3.46) hold for $\ell \leqslant 3$ in all cases, with $\mathcal{Z}_\ell = 0$ except in the case (i).

Step 6: Conclusion. We have shown that (3.45)–(3.46) hold for every ℓ; consequently, (3.33)–(3.34) hold by (3.43)–(3.44).

Moreover, in cases (ii) and (iii), $\mathcal{Z}_\ell = 0$ for $\ell = 1, 2, 3$, and thus $\mathcal{X}_i = a_j r_{ji}\mathcal{Z}_4$; this yields (3.35) by (3.54), which in particular shows that $\mathcal{X}_i > 0$ when $\mathcal{X}_j > 0$ in these cases.

It remains to show $\mathcal{X}_i > 0$ when $\mathcal{X}_j > 0$ in the case (i), i.e., when $\lambda_i > \lambda_j^*$. In this case $\lambda_i^* = \lambda_i$ and $\kappa_i = 0$, and thus by (3.29) and (3.28),

$$\widetilde{X}_i(t) = e^{-\lambda_i t}X_i(t) = \sum_{k=1}^{\infty} e^{-\lambda_i t}\mathbf{1}_{\{T_k\leqslant t\}}Y_k(t-T_k). \tag{3.78}$$

Moreover, if $\mathcal{X}_j > 0$, then $\liminf_{t\to\infty} X_j(t) > 0$ and thus a.s. there is an infinite number of draws of colour j, and thus all T_k are finite. Let K be the (random) smallest k such that $\eta_k > 0$; such k exist a.s. since $\mathbb{E}\,\eta_k = r_{ji} > 0$ by assumption. Then (3.78) implies

$$\widetilde{X}_i(t) \geqslant e^{-\lambda_i T_K}\mathbf{1}_{\{t\geqslant T_K\}}e^{-\lambda_i(t-T_K)}Y_K(t-T_K), \tag{3.79}$$

which a.s. has a strictly positive limit by conditioning on K and applying Lemma 3.3 to Y_K. We have already shown that the limit $\mathcal{X}_i$ in (3.33) exists a.s., and (3.79) then shows $\mathcal{X}_i > 0$ a.s. □

If $\lambda_i \leqslant \lambda_j^*$, then $\mathcal{X}_i$ is determined by $\mathcal{X}_j$, see (3.35). On the contrary, if $\lambda_i > \lambda_j^*$, then $\mathcal{X}_i$ is not determined by $\mathcal{X}_j$, as shown in the following lemma.

Lemma 3.6. *If $\lambda_i > \lambda_j^*$ in Lemma 3.5, then the distribution of $\mathcal{X}_i$ conditioned on $\mathcal{X}_j$ is non-degenerate. In fact, then the conditional distribution $\mathcal{L}(\mathcal{X}_i \mid \mathcal{X}_j)$ is a.s. absolutely continuous.*

Proof. Since $\lambda_i > \lambda_j^*$, we have (3.78). Condition on the entire process $(X_j(t))_{t\geqslant 0}$ and on all T_k and η_k; then the terms in the sum in (3.78) are independent, and, as $t \to \infty$, each term converges a.s. by Lemma 3.3 to a limit that by Lemma 3.4 has an absolutely continuous distribution when $\eta_k > 0$. Since a.s. $\eta_k > 0$ for infinitely many k, it follows that the limit is a.s. (conditionally) absolutely continuous. Since $(X_j(t))_t$ determines $\mathcal{X}_j$, this implies (by Lemma A.1) that the distribution a.s. is absolutely continuous also if we condition only on $\mathcal{X}_j$. □

3.3. **The general case for a single colour.** We have in the preceding subsections considered two special cases. We consider now a single colour i in a general triangular urn. We continue to use the notations (3.29)–(3.30). Recall also that $\mathsf{P}_i := \{j \in \mathsf{Q} : j \to i\}$.

Lemma 3.7. *Let $i \in \mathsf{Q}$, and assume that for every $j \in \mathsf{P}_i$, we have*

$$\widetilde{X}_j(t) \xrightarrow{\text{a.s.}} \mathcal{X}_j \qquad \text{as } t \to \infty, \tag{3.80}$$

for some random variable $\mathcal{X}_j \geqslant 0$, and

$$\big\|\widetilde{X}_j^{**}\big\|_2 < \infty. \tag{3.81}$$

Then

$$\widetilde{X}_i(t) \xrightarrow{\text{a.s.}} \mathcal{X}_i \qquad \text{as } t \to \infty, \tag{3.82}$$

for some $\mathcal{X}_i \geqslant 0$ and

$$\big\|\widetilde{X}_i^{**}\big\|_2 < \infty. \tag{3.83}$$

Furthermore, a.s., if $\mathcal{X}_j > 0$ for every $j \in \mathsf{P}_i$, then $\mathcal{X}_i > 0$.

Proof. If $\mathsf{P}_i = \emptyset$, we are in the situation of Lemma 3.3. Moreover, in this case $X_i(0) > 0$ by our standing assumption (A3). Consequently, in this case the result follows from Lemma 3.3. (Note that $\lambda_i^* = \lambda_i$ and $\kappa_i = 0$ by (2.8) and (2.9).)

In general, we separate the balls of colour i according to their original reason for existing. Formally, we split the colour i and replace it by several "subcolours" (or shades); we define one subcolour labelled i_0, and an additional subcolour i_j for each $j \in \mathsf{P}_i$. These subcolours have the same replacement vector $\boldsymbol{\xi}_i$ as i, with the modification that new balls of colour i always get the same subcolour as the drawn ball. Also, balls of colour i that are added when a ball of some other colour j is drawn get the subcolour i_j. Furthermore, all balls of colour i at time 0 get subcolour i_0. In other words, i_0 is used for descendants of the balls with colour i in the urn at the beginning, and i_j are used for balls of colour i that eventually descend from a ball of colour j. Note that i_0 is minimal, while $\mathsf{P}_{i_j} = \{j\}$.

We thus have

$$X_i(t) = X_{i_0}(t) + \sum_{j \in \mathsf{P}_i} X_{i_j}(t). \tag{3.84}$$

Moreover, in the modified urn with subcolours instead of colour i, the subcolour i_0 is of the type in Section 3.1 (possibly with $x_0 = 0$), and each subcolour i_j ($j \in \mathsf{P}_i$) is of the type in Section 3.2. Hence, Lemmas 3.3 and 3.5 apply, using our assumptions (3.80)–(3.81). It follows from (2.8) that

$$\lambda^*_{i_0} = \lambda_i, \tag{3.85}$$

$$\lambda^*_{i_j} = \lambda_i \vee \lambda^*_j, \qquad j \in \mathsf{P}_i. \tag{3.86}$$

In particular, for every $j \in \mathsf{P}_i \cup \{0\}$, we have $\lambda^*_{i_j} \leqslant \lambda^*_i$. Furthermore, if $\mathsf{P}_i \neq \emptyset$, then

$$\lambda^*_i = \lambda_i \vee \max\{\lambda^*_j : j \in \mathsf{P}_i\} = \max\{\lambda^*_{i_j} : j \in \mathsf{P}_i\}. \tag{3.87}$$

Moreover, (2.9) yields $\kappa_{i_0} = 0$, and implies also that if $\lambda^*_{i_j} = \lambda^*_i$, then $\kappa_{i_j} \leqslant \kappa_i$. Hence, for every $j \in \mathsf{P}_i \cup \{0\}$, and for all large t (at least),

$$t^{-\kappa_i} e^{-\lambda^*_i t} \leqslant t^{-\kappa_{i_j}} e^{-\lambda^*_{i_j} t}. \tag{3.88}$$

Consequently, it follows from (3.84) and the definition (3.29) that, at least for large t,

$$\widetilde{X}_i(t) \leqslant \widetilde{X}_{i_0}(t) + \sum_{j \in \mathsf{P}_i} \widetilde{X}_{i_j}(t), \tag{3.89}$$

and furthermore, using also (3.13) and (3.33),

$$\widetilde{X}_i(t) \xrightarrow{\text{a.s.}} \mathcal{X}_i := \sum_{j \in \mathsf{P}_i \cup \{0\}} \mathcal{X}_{i_j} \mathbf{1}_{\{(\lambda^*_{i_j}, \kappa_{i_j}) = (\lambda^*_i, \kappa_i)\}}. \tag{3.90}$$

This shows (3.82).

Similarly, $(t+1)^{-\kappa_i} e^{-\lambda^*_i t} \leqslant C(t+1)^{-\kappa_{i_j}} e^{-\lambda^*_{i_j} t}$ for every $j \in \mathsf{P}_i \cup \{0\}$ and every $t \geqslant 0$, and thus (3.84) and (3.30) imply

$$\widetilde{X}^{**}_i \leqslant C\widetilde{X}^{**}_{i_0} + C \sum_{j \in \mathsf{P}_i} \widetilde{X}^{**}_{i_j}. \tag{3.91}$$

Hence, (3.83) follows from (3.17) and (3.34).

Finally, assume $\mathcal{X}_j > 0$ for every $j \in \mathsf{P}_i$. If $\mathsf{P}_i = \emptyset$, then, as remarked above, $\mathcal{X}_i > 0$ a.s. by (A3) and Lemma 3.3. On the other hand, if $\mathsf{P}_i \neq \emptyset$, then $\mathcal{X}_{i_j} > 0$ for every $j \in \mathsf{P}_i$ by Lemma 3.5. By (3.87) and (2.9), there exists some $j \in \mathsf{P}_i$ such that $(\lambda^*_{i_j}, \kappa_{i_j}) = (\lambda^*_i, \kappa_i)$, and thus (3.90) implies

$$\mathcal{X}_i \geqslant \mathcal{X}_{i_j} > 0 \tag{3.92}$$

a.s., which completes the proof. □

Remark 3.8. If $\lambda_i > \lambda^*_j$ for every $j \in \mathsf{P}_i$, then $\lambda^*_{i_j} = \lambda_i = \lambda^*_i$ and $\kappa_{i_j} = 0 = \kappa_i$ for every $j \in \mathsf{P}_i \cup \{0\}$, and thus (3.90) yields

$$\mathcal{X}_i = \sum_{j \in \mathsf{P}_i \cup \{0\}} \mathcal{X}_{i_j}. \tag{3.93}$$

On the other hand, suppose that $\lambda_i \leqslant \lambda^*_j$ for some $j \in \mathsf{P}_i$. Then either $\lambda^*_i > \lambda_i = \lambda^*_{i_0}$, or $\lambda_i = \lambda^*_i = \lambda^*_j$ for some $j \in \mathsf{P}_i$, and in the latter case $\kappa_i \geqslant 1 + \kappa_j > \kappa_{i_0} = 0$. Hence, in both cases, $(\lambda^*_{i_0}, \kappa_{i_0}) \neq (\lambda^*_i, \kappa_i)$, and thus the sum in (3.90) is really only over (some) $j \in \mathsf{P}_i$. △

4. The main theorem for continuous time

Theorem 4.1. *Let* $(X_i(t))_{i\in\mathsf{Q}}$ *be a continuous-time triangular Pólya urn satisfying the conditions* (A1)–(A5). *Then, for every colour* $i \in \mathsf{Q}$,

$$t^{-\kappa_i}e^{-\lambda_i^* t}X_i(t) \xrightarrow{\text{a.s.}} \mathcal{X}_i \qquad \text{as } t\to\infty, \tag{4.1}$$

for some random variable $\mathcal{X}_i$ *with* $\mathcal{X}_i > 0$ *a.s.*

Proof. We may choose a total order $<$ of the colours such that (2.5) holds. Taking the colours in this order, we see by Lemma 3.7 and induction that (3.82) and (3.83) hold for every $i \in \mathsf{Q}$, with $\mathcal{X}_i > 0$ a.s. Recalling (3.29), we see that (3.82) is the same as (4.1). ☐

The limits $\mathcal{X}_i$ are strictly positive. We will see in Theorem 6.4 that they are non-degenerate when $\lambda_i^* > 0$; however, in general some of them will be linear combinations of others. We discuss dependencies between the limits in Section 6.

5. Proof of the main theorem for discrete time

Proof of Theorem 1.8. We modify the urn by adding a new colour, 0 say, to Q; let $\mathsf{Q}^+ := \mathsf{Q} \cup \{0\}$ be the new set of colours. All activities a_i and replacements ξ_{ij} with $i,j \in \mathsf{Q}$ remain unchanged. Balls of colour 0 have activity $a_0 := 0$, and thus they are never drawn; in accordance with (A2) we define $\xi_{0j} := 0$ for every $j \in \mathsf{Q}^+$. We further let $\xi_{i0} := 1$ for every $i \in \mathsf{Q}$ with $a_i > 0$. For $i \in \mathsf{Q}$ with $a_i = 0$ we let $\xi_{i0} := 0$, again in accordance with (A2). We let $X_{00} = X_0(0) := 0$, so there are initially no balls of colour 0. Note that the extended urn too satisfies (A1)–(A5).

Since balls of colour 0 never are drawn, we may ignore them and recover the original urn. In other words, the new urn will be the original urn with "dummy balls" of colour 0 added. We may assume that the old and new urn are coupled in this way, which means that for $i \in \mathsf{Q}$, the number of balls of colour i at any time is the same in both urns, so we may unambiguously use the notations X_{ni} and $X_i(t)$ for both urns. Moreover, there are initially no balls of colour 0, but we add exactly 1 ball of colour 0 every time a ball is drawn. Hence, in the discrete-time version, the number of dummy balls at time n equals n; thus $X_{n0} = n$. In the continuous-time model, $X_0(t)$ equals the number of draws up to time t , and in particular, cf. (2.19),

$$X_0(\widehat{T}_n) = X_{n0} = n. \tag{5.1}$$

The new urn is obviously also triangular, with $i \to 0$ for every $i \in \mathsf{Q}$ with $a_i > 0$. We have $\lambda_0 = 0$. Since every $\lambda_i \geqslant 0$, and we have $i \to 0$ when $\lambda_i > 0$ (and thus $a_i > 0$ by (A2)), it follows from (2.8) and (2.10) that

$$\lambda_0^* = \max\{\lambda_j : j \in \mathsf{Q}\} = \widehat{\lambda}. \tag{5.2}$$

If $\widehat{\lambda} > 0$, then is further easily seen from (2.9) and (2.11) that

$$\kappa_0 = \max\{\kappa_i : i \in \mathsf{Q} \text{ and } \lambda_i^* = \widehat{\lambda}\} = \widehat{\kappa}. \tag{5.3}$$

If $\widehat{\lambda} = 0$, which means that $\lambda_i = 0$ for every i, we have by (2.9) and (2.12) instead

$$\begin{aligned}\kappa_0 &= \max\bigl\{\kappa : \exists i_1 \prec i_2 \prec \cdots \prec i_{\kappa+1} = 0 \text{ in } \mathsf{Q}^+\bigr\}\\ &= \max\bigl\{\kappa : \exists i_1 \prec i_2 \prec \cdots \prec i_{\kappa} \text{ with } a_{i_\kappa} > 0 \text{ in } \mathsf{Q}\bigr\}\\ &= 1 + \max\bigl\{\kappa_j : j \in \mathsf{Q} \text{ with } a_j > 0\bigr\}\\ &= \widehat{\kappa}_0.\end{aligned} \tag{5.4}$$

We apply Theorem 4.1 to the new urn. In particular, taking $i = 0$ in (4.1) yields, using (5.1) and (5.2)–(5.4) and recalling that $\widehat{T}_n \to \infty$ a.s.,

$$\widehat{T}_n^{-\kappa_0} e^{-\widehat{\lambda}\widehat{T}_n} n = \widehat{T}_n^{-\kappa_0} e^{-\widehat{\lambda}\widehat{T}_n} X_0(\widehat{T}_n) \xrightarrow{\text{a.s.}} \mathcal{X}_0, \qquad \text{as } n \to \infty. \tag{5.5}$$

Recall that $0 < \mathcal{X}_0 < \infty$ a.s.

Case 1: $\widehat{\lambda} > 0$. Consider first the case $\widehat{\lambda} > 0$; then $\kappa_0 = \widehat{\kappa}$ by (5.3). It follows from (5.5) that a.s., as $n \to \infty$,

$$\log n - \widehat{\lambda}\widehat{T}_n - \widehat{\kappa} \log \widehat{T}_n \to \log \mathcal{X}_0 \tag{5.6}$$

which yields, since $\widehat{T}_n \to \infty$,

$$\frac{\log n}{\widehat{T}_n} \to \widehat{\lambda}, \tag{5.7}$$

and thus

$$\log\log n - \log \widehat{T}_n \to \log \widehat{\lambda}. \tag{5.8}$$

Using (5.8) in (5.6) yields, a.s.,

$$\widehat{T}_n = \frac{1}{\widehat{\lambda}}\bigl(\log n - \widehat{\kappa} \log\log n + \widehat{\kappa} \log \widehat{\lambda} - \log \mathcal{X}_0 + o(1)\bigr). \tag{5.9}$$

Now let $i \in \mathsf{Q}$. Then (5.9) yields, as $n \to \infty$, a.s.,

$$\widehat{T}_n^{\kappa_i} e^{\lambda_i^* \widehat{T}_n} \sim \widehat{\lambda}^{-\kappa_i + \widehat{\kappa}\lambda_i^*/\widehat{\lambda}} \mathcal{X}_0^{-\lambda_i^*/\widehat{\lambda}} n^{\lambda_i^*/\widehat{\lambda}} (\log n)^{\kappa_i - \widehat{\kappa}\lambda_i^*/\widehat{\lambda}}. \tag{5.10}$$

Consequently, taking $t = \widehat{T}_n$ in (4.1) yields, as $n \to \infty$, using the notation (2.13),

$$\frac{X_{ni}}{n^{\lambda_i^*/\widehat{\lambda}} \log^{\gamma_i} n} = \frac{X_i(\widehat{T}_n)}{n^{\lambda_i^*/\widehat{\lambda}} (\log n)^{\kappa_i - \widehat{\kappa}\lambda_i^*/\widehat{\lambda}}} \xrightarrow{\text{a.s.}} \widehat{\mathcal{X}}_i \tag{5.11}$$

with

$$\widehat{\mathcal{X}}_i := \widehat{\lambda}^{-\kappa_i + \widehat{\kappa}\lambda_i^*/\widehat{\lambda}} \mathcal{X}_0^{-\lambda_i^*/\widehat{\lambda}} \mathcal{X}_i = \widehat{\lambda}^{-\gamma_i} \mathcal{X}_0^{-\lambda_i^*/\widehat{\lambda}} \mathcal{X}_i. \tag{5.12}$$

This shows (1.3).

Case 2: $\widehat{\lambda} = 0$. In the case $\widehat{\lambda} = 0$, we have instead $\kappa_0 = \widehat{\kappa}_0$ by (5.4). Moreover, (5.5) now yields that as $n \to \infty$, a.s.,

$$\widehat{T}_n \sim \mathcal{X}_0^{-1/\widehat{\kappa}_0} n^{1/\widehat{\kappa}_0} \tag{5.13}$$

and thus (4.1) yields

$$\frac{X_{ni}}{n^{\kappa_i/\widehat{\kappa}_0}} = \frac{X_i(\widehat{T}_n)}{n^{\kappa_i/\widehat{\kappa}_0}} \xrightarrow{\text{a.s.}} \widehat{\mathcal{X}}_i := \mathcal{X}_0^{-\kappa_i/\widehat{\kappa}_0} \mathcal{X}_i. \tag{5.14}$$

This shows (1.4), and completes the proof of Theorem 1.8. □

The limits $\widehat{\mathcal{X}}_i$ are all strictly positive. We return to the question whether they are degenerate in Section 7.

Remark 5.1. Suppose that $\widehat{\lambda} > 0$, so that (5.11) holds. As a sanity check, we note that (5.11) and (5.12) imply, recalling (2.10) and (2.11), that if we define

$$\mathsf{Q}_* := \{i \in \mathsf{Q} : \lambda_i^* = \widehat{\lambda} \text{ and } \kappa_i = \widehat{\kappa}\}, \tag{5.15}$$

which is nonempty, then

$$\frac{X_{ni}}{n} \xrightarrow{\text{a.s.}} \begin{cases} \widehat{\mathcal{X}}_i = \mathcal{X}_i/\mathcal{X}_0 & \text{if } i \in \mathsf{Q}_*, \\ 0 & \text{otherwise.} \end{cases} \tag{5.16}$$

Hence, the total number of balls grows linearly, as should be expected; moreover, the distribution of colours in the urn is asymptotically concentrated on Q_*.

Note also that (3.90) for the dummy colour 0 yields, recalling (5.2)–(5.3) and using (3.35),

$$\mathcal{X}_0 = \sum_{j \in \mathsf{Q}_*} \mathcal{X}_{0_j} = \sum_{j \in \mathsf{Q}_*} \frac{a_j}{\lambda_j^*} \mathcal{X}_j = \widehat{\lambda}^{-1} \sum_{j \in \mathsf{Q}_*} a_j \mathcal{X}_j. \tag{5.17}$$

By (5.16) and (5.17), we have for the total activity in the urn

$$\frac{1}{n} \sum_{i \in \mathsf{Q}} a_i X_{ni} \xrightarrow{\text{a.s.}} \frac{1}{\mathcal{X}_0} \sum_{i \in \mathsf{Q}_*} a_i \mathcal{X}_i = \widehat{\lambda}. \tag{5.18}$$

△

Remark 5.2. If we instead suppose $\widehat{\lambda} = 0$, which implies $\lambda_i = \lambda_i^* = 0$ for every i, then (5.14) similarly shows that if we now define

$$\mathsf{Q}_* := \{i : \kappa_i = \widehat{\kappa}_0\}, \tag{5.19}$$

then (5.16) still holds. Note that it is quite possible that Q_* is empty; this happens precisely when $\widehat{\kappa}_0 > \widehat{\kappa}$, which by (5.4) happens when some i with $\kappa_i = \widehat{\kappa}$ has activity $a_i > 0$. (Such a colour cannot have any descendants, since all colours have the same $\lambda_j^* = 0$; hence we must have $\boldsymbol{\xi}_i = 0$ a.s., which means that such colours will be drawn, but nothing happens to the urn at these draws.) In this case the total number of balls is a.s. $o(n)$, and the colour distribution is asymptotically concentrated on the colours i such that $\kappa_i = \widehat{\kappa} = \widehat{\kappa}_0 - 1$.

On the other hand, if $\mathsf{Q}_* \neq \emptyset$, which means that $\widehat{\kappa}_0 = \widehat{\kappa}$, and thus by (5.4) $a_i = 0$ for every i with $\kappa_i = \widehat{\kappa}$, let $\mathsf{Q}_{*-} := \{j : \kappa_j = \widehat{\kappa} - 1\}$. Then (3.90) and (3.35) yield, since $j \to 0$ if and only if $a_j > 0$, and then $\kappa_{0_j} = \kappa_j + 1$,

$$\mathcal{X}_0 = \sum_{j \in \mathsf{P}_0 \cap \mathsf{Q}_{*-}} \mathcal{X}_{0_j} = \sum_{j \in \mathsf{Q}_{*-}} \frac{a_j}{\widehat{\kappa}} \mathcal{X}_j = \widehat{\kappa}^{-1} \sum_{j \in \mathsf{Q}_{*-}} a_j \mathcal{X}_j. \tag{5.20}$$

△

6. Dependencies between the limits

The limits $\mathcal{X}_i$ in Theorem 4.1 are non-degenerate except in extreme cases, as shown below, but there are frequently linear dependencies between them. To explore this, we introduce more terminology.

We say that a colour i is a *leader* if $\lambda_j < \lambda_i$ for every $j \prec i$, and a *follower* otherwise. (In particular, a minimal colour i is a leader.) We have, recalling (2.8),

$$i \text{ is a leader} \iff \Bigl(j \prec i \implies \lambda_j < \lambda_i\Bigr) \iff \Bigl(j \to i \implies \lambda_j^* < \lambda_i\Bigr), \tag{6.1}$$

$$i \text{ is a follower} \iff \Bigl(\exists j \prec i \text{ with } \lambda_j \geqslant \lambda_i\Bigr) \iff \Bigl(\exists j \to i \text{ with } \lambda^*_j \geqslant \lambda_i\Bigr). \qquad (6.2)$$

By (2.8)–(2.9),

$$i \text{ is a leader} \iff \lambda^*_i = \lambda_i \text{ and } \kappa_i = 0. \qquad (6.3)$$

If i is a follower, let A_i be the set of ancestors of i that have maximal λ_j, i.e., $\lambda_j = \lambda^*_i$, and such that furthermore there is a path from j to i with the maximum number $\kappa_i + 1$ of colours ℓ with $\lambda_\ell = \lambda^*_i$. (Recall (2.8) and (2.9).) Thus

$$\mathsf{A}_i := \{j \prec i : \exists i_1 = j \prec i_2 \prec \cdots \prec i_{\kappa_i+1} \preceq i \text{ with } \lambda_{i_1} = \cdots = \lambda_{i_{\kappa+1}} = \lambda^*_i\}. \qquad (6.4)$$

Note that A_i is non-empty for every follower i. Moreover, it is easily seen from (6.1)–(6.2) and (6.4) that every $j \in \mathsf{A}_i$ is a leader. We may say that i follows the leaders in A_i; note that a follower may follow several leaders. For completeness, we define $\mathsf{A}_i := \{i\}$ when i is a leader. Note that, by (6.3) and (6.4),

$$\lambda_j = \lambda^*_j = \lambda^*_i \qquad \text{for every } j \in \mathsf{A}_i. \qquad (6.5)$$

We show first that the variables $\mathcal{X}_i$ are determined by the ones for leaders i.

Lemma 6.1. *If $i \in \mathsf{Q}$, then $\mathcal{X}_i$ is a linear combination*

$$\mathcal{X}_i = \sum_{k \in \mathsf{A}_i} c_{ik} \mathcal{X}_k \qquad (6.6)$$

with strictly positive coefficients c_{ik}.

Proof. Note first that (6.6) is trivial when i is a leader, with $c_{ii} = 1$. We may thus suppose that i is a follower.

By induction on the colour i, we may assume that the formula (6.6) holds for every colour that is an ancestor of i, and in particular for every $j \in \mathsf{P}_i$.

By Remark 3.8, we only have to consider $j \in \mathsf{P}_i$ in (3.90). Let

$$\mathsf{P}'_i := \bigl\{j \in \mathsf{P}_i : (\lambda^*_{i_j}, \kappa_{i_j}) = (\lambda^*_i, \kappa_i)\bigr\}, \qquad (6.7)$$

so that the sum in (3.90) really is over $j \in \mathsf{P}'_i$. By (6.3), since we now assume that i is a follower, either $\lambda^*_i > \lambda_i$ or $\kappa_i > 0$ (or both). Hence, if $j \in \mathsf{P}'_i$, then either $\lambda^*_{i_j} = \lambda^*_i > \lambda_i$ or $\kappa_{i_j} = \kappa_i > 0$; since $\lambda^*_{i_j} = \lambda_i \vee \lambda^*_j$, it follows in both cases (using (2.9) in the latter) that $\lambda^*_j = \lambda^*_{i_j} \geqslant \lambda_i$ and thus by (3.35)

$$\mathcal{X}_{i_j} = c'_{ij} \mathcal{X}_j \qquad (6.8)$$

for some constant $c'_{ij} > 0$. Moreover, we have $\lambda^*_j = \lambda^*_j \vee \lambda_i = \lambda^*_{i_j} = \lambda^*_i$ and it follows from (6.7) and (2.9) that

$$\kappa_i = \kappa_{i_j} = \kappa_j + \mathbf{1}_{\{\lambda_i = \lambda^*_i\}}. \qquad (6.9)$$

If $j \in \mathsf{P}'_i$ is a leader, so $\lambda_j = \lambda^*_j = \lambda^*_i$ and $\kappa_j = 0$ by (6.3), it follows from (6.4) and (6.9) that $j \in \mathsf{A}_i$. Similarly, if $j \in \mathsf{P}'_i$ is a follower, then it follows from (6.4) and (6.9) that if $k \in \mathsf{A}_j$, then a chain from k to j of the (maximal) type in (6.4) can be extended to a chain from k to i of the same type, and thus $k \in \mathsf{A}_i$; consequently, if $j \in \mathsf{P}'_i$ is a follower, then $\mathsf{A}_j \subseteq \mathsf{A}_i$. The result (6.6), with some $c_{ij} \geqslant 0$, now follows by (3.90), (6.8), and induction.

Finally, if $k \in \mathsf{A}_i$, let $i_1 = k, \dots, i_{\kappa_i+1} \preceq i$ be as in (6.4), choose a path in Q from k to i that contains all i_ℓ, and let j be the last colour in this path before i. Then it follows from (6.4) that $k \in \mathsf{A}_j$, and thus by induction $c_{jk} > 0$, which implies $c_{ik} \geqslant c'_{ij} c_{jk} > 0$. ∎

Motivated by (6.6), we turn to considering $\mathcal{X}_i$ for leaders i. We first consider a trivial exceptional case.

Lemma 6.2. *Let i be a colour with $\lambda_i = 0$. Then i is a leader if and only if i is minimal, i.e., there is no colour j with $j \to i$. In this case $\mathcal{X}_i = X_i(0)$ is a deterministic positive constant.*

Proof. We have already remarked that a minimal colour is a leader. Conversely, if i is a leader with $\lambda_i = 0$, then (6.1) shows that $j \to i$ is impossible; hence i is minimal. In this case, no balls of colour i are added by draws of balls of other colours. Furthermore, since $\lambda_i = 0$, also no balls of colour i are added by draws of colour i. Consequently, $X_i(t) = X_i(0)$ for all $t \geqslant 0$, and (3.80) holds trivially with $\widetilde{X}_i(t) = X_i(t)$ and $\mathcal{X}_i = X_i(0)$. □

We next show that except for the case in Lemma 6.2, the distribution of $\mathcal{X}_i$ for a leader i is absolutely continuous, and thus non-degenerate.

Lemma 6.3. (i) *Let i be a leader with $\lambda_i > 0$, and let E be a set of colours that contains neither i nor any descendant of i, i.e., if $j \in \mathsf{E}$ then $j \nsucceq i$. Then the conditional distribution $\mathcal{L}(\mathcal{X}_i \mid \mathcal{X}_j, j \in \mathsf{E})$ is a.s. absolutely continuous. In particular, the distribution of $\mathcal{X}_i$ is absolutely continuous.*

(ii) *More generally, let L be a (non-empty) set of leaders with $\lambda_i > 0$ for every $i \in \mathsf{L}$, and suppose that L is an anti-chain in Q, i.e., $i \nprec j$ when $i, j \in \mathsf{L}$. Let further E be a set of colours such that if $j \in \mathsf{E}$, then $j \nsucceq i$ for every $i \in \mathsf{L}$. Then the joint conditional distribution $\mathcal{L}((\mathcal{X}_i)_{i\in\mathsf{L}} \mid \mathcal{X}_j, j \in \mathsf{E})$ is a.s. absolutely continuous. (This is a distribution in $\mathbb{R}^{|\mathsf{L}|}$.)*

Proof. It suffices to consider (ii). The conclusion (conditional absolute continuity) is preserved if we reduce E to a smaller set, see Lemma A.1 and Remark A.2. Thus we may assume that E is maximal, i.e.,

$$\mathsf{E} = \{j \in \mathsf{Q} : j \nsucceq i, \forall i \in \mathsf{L}\}. \tag{6.10}$$

Note that, if $i \in \mathsf{L}$ and $j \prec i$, then we cannot have $j \succeq k$ for any $k \in \mathsf{L}$, since this would imply $k \prec i$, contradicting the assumption that L is an antichain. Consequently, by (6.10), if $i \in \mathsf{L}$ and $j \prec i$, then $j \in \mathsf{E}$.

We argue as in the proof of Lemma 3.6, and condition on the entire processes $(X_j(t))_{t\geqslant 0}$, $j \in \mathsf{E}$, and also on the times of all draws of a color $j \in \mathsf{E}$, and on all replacement vectors $\boldsymbol{\xi}_j^{(n)}$ ($n \geqslant 1$, $j \in \mathsf{E}$).

Let $i \in \mathsf{L}$. If $j \in \mathsf{P}_i$, so $j \to i$, then $\lambda_j^* < \lambda_i = \lambda_i^*$ by the assumption that i is a leader. We use again the decomposition (3.84), and note that each $e^{-\lambda_i t} X_{i_j}(t)$ ($j \in \mathsf{P}_i$) can be written as a sum as in (3.78) with (conditionally) independent terms, and it follows as in the proof of Lemma 3.6 that each $\mathcal{X}_{i_j}$ has (conditionally) an absolutely continuous distribution. Furthermore, $\mathcal{X}_{i_0}$ is independent of conditioning on colours $j \in \mathsf{E}$, and if $X_i(0) > 0$, then its distribution too is absolutely continuous by Lemma 3.4 (since we assume $\lambda_i > 0$). Moreover (still conditionally), all $\mathcal{X}_{i_j}$ and $\mathcal{X}_{i_0}$ ($i \in \mathsf{L}$, $j \in \mathsf{P}_i$) are independent.

Since i is a leader, Remark 3.8 shows that (3.93) holds. We have shown that (conditionally) the terms in the sum in (3.93) are independent, and all are absolutely continuous except $\mathcal{X}_{i_0}$ when $X_i(0) = 0$. Thus our assumption (A3) implies that the sum contains at least one absolutely continuous summand, and hence $\mathcal{X}_i$ is (conditionally) absolutely continuous.

Moreover, this argument shows that $\mathcal{X}_i$ for different $i \in \mathsf{L}$ are (conditionally) independent, and since each has an absolutely continuous distribution, their joint distribution is (conditionally) absolutely continuous.

Finally, Lemma A.1 (again with Remark A.2) shows that the same holds if we instead condition on $\mathcal{X}_j$, $j \in \mathsf{E}$. ☐

Theorem 6.4. *Let L be a set of leaders such that $\lambda_i > 0$ for every $i \in \mathsf{L}$. Then the joint distribution $\mathcal{L}\big((\mathcal{X}_i)_{i\in\mathsf{L}}\big)$ is absolutely continuous.*

Proof. Order the elements of the set $\{\lambda_i : i \in \mathsf{L}\}$ as $\lambda_{(1)} < \cdots < \lambda_{(r)}$. Let

$$\mathsf{L}_\ell := \{i \in \mathsf{L} : \lambda_i = \lambda_{(\ell)}\}, \qquad 1 \leqslant \ell \leqslant r, \tag{6.11}$$

$$\mathsf{L}_{\leqslant\ell} := \bigcup_{k:=1}^{\ell} \mathsf{L}_k, \qquad 0 \leqslant \ell \leqslant r. \tag{6.12}$$

(Thus, $\mathsf{L}_{\leqslant 0} = \varnothing$.) Let $\ell \in \{1, \dots, r\}$. If $i, j \in \mathsf{L}_\ell$, then (6.1) shows that $j \not\prec i$, and thus L_ℓ is an antichain. Furthermore, if $i \in \mathsf{L}_\ell$ and $j \in \mathsf{L}_{\leqslant\ell-1}$, then $\lambda_j^* = \lambda_j < \lambda_i$, and thus $i \not\preccurlyeq j$. Consequently, Lemma 6.3(ii) shows that the conditional distribution of $(\mathcal{X}_i)_{i\in\mathsf{L}_\ell}$ given $(\mathcal{X}_j)_{j\in\mathsf{L}_{\leqslant\ell-1}}$ is a.s. absolutely continuous.

It now follows from Lemma A.3 by induction that the distribution of $(\mathcal{X}_i)_{i\in\mathsf{L}_{\leqslant\ell}}$ is absolutely continuous for $\ell = 1, \dots, r$. Taking $\ell = r$ yields the result, since $\mathsf{L}_r = \mathsf{L}$. ☐

Theorem 6.5. *Let $i \in \mathsf{Q}$ be any colour.*

(i) *If $\lambda_i^* = 0$, then $\mathcal{X}_i$ is a deterministic positive constant.*
(ii) *If $\lambda_i^* > 0$, then $\mathcal{X}_i$ has an absolutely continuous distribution.*

Proof. (i): Every $k \in \mathsf{A}_i$ is a leader with $\lambda_k = \lambda_i^* = 0$. Thus Lemma 6.2 shows that $\mathcal{X}_k$ is deterministic for $k \in \mathsf{A}_i$. Consequently, (6.6) shows that $\mathcal{X}_i$ too is deterministic.

(ii): Every $k \in \mathsf{A}_i$ is a leader with $\lambda_k = \lambda_i^* > 0$. Thus Theorem 6.4 shows that the joint distribution of $(\mathcal{X}_k)_{k\in\mathsf{A}_i}$ is absolutely continuous. It follows from (6.6) that the distribution of $\mathcal{X}_i$ is absolutely continuous. (See Lemma A.4.) ☐

Corollary 6.6. *If $\lambda_i^* > 0$, then the coefficients c_{ik} in Lemma 6.1 are uniquely determined.*

Proof. Suppose, on the contrary, that (6.6) holds a.s. for two different sets of coefficients $(c_{ik})_k$ and $(c'_{ik})_k$. Let $b_k := c_{ik} - c'_{ik}$. Then $\sum_{k\in\mathsf{A}_i} b_k\mathcal{X}_k = 0$ a.s., and thus $(\mathcal{X}_k)_{k\in\mathsf{A}_i}$ a.s. lies in the certain hyperplane orthogonal to (b_k). However, this contradicts Theorem 6.5 which says that the distribution of $(\mathcal{X}_k)_{k\in\mathsf{A}_i}$ is absolutely continuous. (See also Lemma A.4.) This contradiction proves the claim. ☐

The coefficients c_{ik} in (6.6) can be found by the recursive procedure in the proof of Lemma 6.1. We proceed to show that they also can be found as eigenvectors of suitable submatrices of the weighted mean replacement matrix $(a_i r_{ij})_{i,j\in\mathsf{Q}}$. In the sequel, we let c_{ik} be the coefficient given by the inductive proof of Lemma 6.1; this determines c_{ik} uniquely for $k \in \mathsf{A}_i$ also when $\lambda_i^* = 0$. We further define $c_{ik} := 0$ if $k \notin \mathsf{A}_i$.

Let ν be a leader, and let

$$\mathsf{D}_\nu := \{i : \lambda_i^* = \lambda_\nu\}, \tag{6.13}$$

$$\mathsf{D}_\nu^\kappa := \{i \in \mathsf{D}_\nu : \kappa_i = \kappa\}, \qquad \kappa = 0, 1, \dots. \tag{6.14}$$

(The sets D_ν^κ are empty from some κ on; we only consider κ with $\mathsf{D}_\nu^\kappa \neq \emptyset$. Note also that $\mathsf{D}_\nu = \mathsf{D}_{\nu'}$ if ν' is another leader with $\lambda_{\nu'} = \lambda_\nu$.) Note that if $c_{i\nu} > 0$, then $\nu \in \mathsf{A}_i$, and thus $i \in \mathsf{D}_\nu$.

Let $i \in \mathsf{D}_\nu^\kappa$, and suppose first that $\lambda_i < \lambda_i^* = \lambda_\nu$. Then it follows from Remark 3.8 that in (3.90), we only have to sum over $j \in \mathsf{P}_i$. Moreover, if $j \in \mathsf{P}_i$ and $\lambda_{i_j}^* = \lambda_i^*$, then $\lambda_j^* = \lambda_{i_j}^* = \lambda_i^*$ and $\kappa_{i_j} = \kappa_j$. Hence, in (3.90) we only have to sum over $j \in \mathsf{P}_i \cap \mathsf{D}_\nu^\kappa$. Since $j \in \mathsf{P}_i \iff j \to i \iff r_{ji} > 0$ and $j \neq i$ by (2.3), it follows from (3.90) and (3.35) that

$$\mathcal{X}_i = \sum_{j \in \mathsf{D}_\nu^\kappa \cap \mathsf{P}_i} \mathcal{X}_{i_j} = \sum_{j \in \mathsf{D}_\nu^\kappa \setminus \{i\}} \frac{a_j r_{ji}}{\lambda_j^* - \lambda_i} \mathcal{X}_j. \tag{6.15}$$

Consequently, the coefficient $c_{i\nu}$ in (6.6) is given by

$$c_{i\nu} = \sum_{j \in \mathsf{D}_\nu^\kappa \setminus \{i\}} \frac{a_j r_{ji}}{\lambda_j^* - \lambda_i} c_{j\nu}. \tag{6.16}$$

Since $j \in \mathsf{D}_\nu^\kappa$ implies $\lambda_j^* = \lambda_\nu$, (6.16) yields

$$(\lambda_\nu - \lambda_i) c_{i\nu} = \sum_{j \in \mathsf{D}_\nu^\kappa \setminus \{i\}} a_j r_{ji} c_{j\nu}, \tag{6.17}$$

for every $i \in \mathsf{D}_\nu^\kappa$ with $\lambda_i < \lambda_i^*$.

On the other hand, suppose that $i \in \mathsf{D}_\nu^\kappa$ with $\lambda_i = \lambda_i^* = \lambda_\nu$. If $j \in \mathsf{D}_\nu \setminus \{i\}$ and $r_{ji} > 0$, then $\lambda_j^* = \lambda_\nu = \lambda_i^*$ and $j \to i$, and thus $\kappa_i \geqslant \kappa_j + 1$, since a maximal chain in the definition (2.9) of κ_j can be extended by i. Hence, if $j \in \mathsf{D}_\nu^\kappa \setminus \{i\}$, then $r_{ji} = 0$. It follows that (6.17) holds trivially for $i \in \mathsf{D}_\nu^\kappa$ with $\lambda_i = \lambda_i^*$.

Consequently, (6.17) holds for every $i \in \mathsf{D}_\nu^\kappa$, and thus, recalling $\lambda_i = a_i r_{ii}$,

$$\sum_{j \in \mathsf{D}_\nu^\kappa} a_j r_{ji} c_{j\nu} = (\lambda_\nu - \lambda_i) c_{i\nu} + a_i r_{ii} c_{i\nu} = \lambda_\nu c_{i\nu}, \qquad i \in \mathsf{D}_\nu^\kappa. \tag{6.18}$$

We summarize, and elaborate, this as a lemma. We say that $i \in \mathsf{Q}$ is a *subleader* if $\lambda_i = \lambda_i^*$ but i is not a leader; recalling (6.3) we thus have

$$i \text{ is a subleader} \iff \lambda_i^* = \lambda_i \text{ and } \kappa_i \geqslant 1. \tag{6.19}$$

Lemma 6.7. *For any leader ν, and any $\kappa \geqslant 0$ such that $D_\nu^\kappa \neq \emptyset$, we have (6.18), and thus $(c_{i\nu})_{i \in \mathsf{D}_\nu^\kappa}$ is a left eigenvector of the triangular matrix $(a_i r_{ij})_{i,j \in \mathsf{D}_\nu^\kappa}$ for its largest eigenvalue λ_ν.*

For $\kappa = 0$, the value $c_{i\nu}$ for a leader $i \in \mathsf{D}_\nu^0$ is determined by $c_{i\nu} = \delta_{i\nu}$ (i.e., 1 when $i = \nu$ and 0 otherwise), and these values for the leaders determine the eigenvector $(c_{i\nu})_{i \in \mathsf{D}_\nu^\kappa}$ uniquely.

For $\kappa \geqslant 1$, the value $c_{i\nu}$ for a subleader $i \in \mathsf{D}_\nu^\kappa$ is determined recursively from the values $c_{j\nu}$ for $j \in \mathsf{D}_\nu^{\kappa-1}$ by

$$c_{i\nu} = \sum_{j \in \mathsf{D}_\nu^{\kappa-1}} \frac{a_j r_{ji}}{\kappa} c_{j\nu}, \tag{6.20}$$

and these values for the subleaders determine the eigenvector $(c_{i\nu})_{i \in \mathsf{D}_\nu^\kappa}$ uniquely.

Proof. We have shown that (6.18) holds, and thus $(c_{i\nu})_{i \in \mathsf{D}_\nu^\kappa}$ is a left eigenvector.

Since the matrix $A_\kappa := (a_j r_{ij})_{i,j \in \mathsf{D}_\nu^\kappa}$ is triangular, its eigenvalues are its diagonal elements $a_j r_{jj} = \lambda_j$, $j \in \mathsf{D}_\nu^\kappa$. The definition (2.9) of κ_i implies that if $\mathsf{D}_\nu^\kappa \neq \emptyset$, then there exists $i \in \mathsf{D}_\nu^\kappa$ with $\lambda_i = \lambda_i^*$, i.e. a leader (if $\kappa = 0$) or a subleader (if $\kappa \geqslant 1$).

Since $\lambda_i \leqslant \lambda_i^* = \lambda_\nu$ for every $i \in \mathsf{D}_\nu^\kappa$, it follows that λ_ν^* is the largest eigenvalue, and that its multiplicity equals the number of leaders or subleaders in D_ν^κ.

Let L_κ be the set of leaders or subleaders in D_ν^κ, and consider the projection $\Pi_\kappa : \mathbb{R}^{\mathsf{D}_\nu^\kappa} \to \mathbb{R}^{\mathsf{L}_\kappa}$ mapping $(x_i)_{i\in\mathsf{D}_\nu^\kappa} \mapsto (x_i)_{i\in\mathsf{L}_\kappa}$. Let V_κ be the left eigenspace of the matrix A_κ for its largest eigenvalue λ_ν. The recursion in the proof of Lemma 6.1 and the calculations (6.15)–(6.17) show, more generally, that any vector $(x_i)_{i\in\mathsf{L}_\kappa}$ can be extended to a vector $(x_i)_{i\in\mathsf{D}_\nu^\kappa}$ that belongs to V_κ. In other words, the projection Π_κ maps V_κ onto $\mathbb{R}^{\mathsf{L}_\kappa}$. These spaces have the same dimension $|\mathsf{L}_\kappa|$, and thus $\Pi_\kappa : V_\kappa \to \mathbb{R}^{\mathsf{L}_\kappa}$ is a bijection. Consequently, an eigenvector is determined by its values for leaders or subleaders.

It is trivial that $c_{i\nu} = \delta_{i\nu}$ for a leader i.

For $\kappa \geqslant 1$, let i be a subleader in D_ν^κ, so $\lambda_i = \lambda_i^*$. Since $\kappa_i = \kappa > 0$, it follows from Remark 3.8 that $\lambda_i = \lambda_j^*$ for some $j \in \mathsf{P}_i$, and furthermore that $\kappa_j + 1 \leqslant \kappa_i$ for every such j. Since then also $\lambda_{i_j}^* = \lambda_j^* = \lambda_i^*$ and $\kappa_{i_j} = \kappa_j + 1$, it follows that in (3.90) we only have to sum over $j \in \mathsf{D}_\nu^{\kappa-1}$, which together with (3.35) yields (6.20). □

Remark 6.8. It is well known that the eigenvalues and eigenvectors of the weighted mean replacement matrix $(a_i r_{ij})_{i,j\in\mathsf{Q}}$ are important for the asymptotics of Pólya urns in general; see for example [4, Section V.9.3] and [27] for irreducible urns. Lemma 6.7, where we consider eigenvectors of certain submatrices, is inspired by a special case in [12], see Example 14.12. △

7. Degenerate limits in discrete time

Consider now the limits $\widehat{\mathcal{X}}_i$ in Theorem 1.8 for the discrete-time urn.

Theorem 7.1. *Let $i \in \mathsf{Q}$.*

(i) *If $\lambda_i^* = 0$, then $\widehat{\mathcal{X}}_i$ is a positive constant.*

(ii) *If $0 < \lambda_i^* < \widehat{\lambda}$, then $\widehat{\mathcal{X}}_i$ has an absolutely continuous distribution.*

(iii) *If $\lambda_i^* = \widehat{\lambda} > 0$, then $\widehat{\mathcal{X}}_i$ is either constant or has an absolutely continuous distribution; both alternatives are possible.*

Proof. Consider the urn with an added dummy colour 0 as in Section 5.

(i): In this case, Theorem 6.5 shows that $\mathcal{X}_i$ is a positive constant. If $\widehat{\lambda} > 0$, then (5.12) shows that $\widehat{\mathcal{X}}_i = \widehat{\lambda}^{-\kappa_i}\mathcal{X}_i$ is a constant. If $\widehat{\lambda} = 0$, then (5.2) and Theorem 6.5 show that also $\mathcal{X}_0$ is a positive constant; thus (5.14) shows that $\widehat{\mathcal{X}}_i$ is a constant.

(ii): By (5.2), we have $\lambda_0^* = \widehat{\lambda} > 0$. By Lemma 6.1, $\mathcal{X}_i$ is a linear combination (6.6) of $\mathcal{X}_k$ for leaders k with $\lambda_k^* = \lambda_i^*$, and $\mathcal{X}_0$ is a similar linear combination, now for leaders k with $\lambda_k^* = \lambda_0^* = \widehat{\lambda}$. The two sets of leaders are disjoint, and thus Theorem 6.4 implies, using Lemma A.4, that the distribution of $(\mathcal{X}_i, \mathcal{X}_0)$ is absolutely continuous in $\mathbb{R}^2$. It is then easily seen from (5.12) that the distribution of $\widehat{\mathcal{X}}_i$ is absolutely continuous.

(iii): In this case, Lemma 6.1 shows that both $\mathcal{X}_i$ and $\mathcal{X}_0$ are linear combinations (6.6) of $\mathcal{X}_k$ for leaders k with $\lambda_k^* = \widehat{\lambda}$. If the two vectors of coefficients of these linear combinations are proportional, then $\mathcal{X}_i$ and $\mathcal{X}_0$ are proportional; since now (5.12) shows that $\widehat{\mathcal{X}} = c\mathcal{X}_i/\mathcal{X}_0$ for some constant c, it follows that $\widehat{\mathcal{X}}$ is constant. On the other hand, if the vectors of coefficients are not proportional, then Lemma A.4 shows that the distribution of $(\mathcal{X}_i, \mathcal{X}_0)$ is absolutely continuous in $\mathbb{R}^2$, and as in (ii), it follows easily from (5.12) that the distribution of $\widehat{\mathcal{X}}_i$ is absolutely continuous. □

An example where $\lambda_i^* = \widehat{\lambda}$ and $\widehat{\mathcal{X}}_i$ is absolutely continuous is given by the classical Pólya urn, see Example 14.1. Many examples with degenerate $\widehat{\mathcal{X}}_i$ are provided by the following theorem; see also the examples in Section 14.

Theorem 7.2. *Suppose that there is only one leader ν with $\lambda_\nu = \widehat{\lambda}$. (This is equivalent to $\lambda_\nu = \widehat{\lambda}$ and that $i \succeq \nu$ for every colour i with $\lambda_i = \widehat{\lambda}$.) Then $\widehat{\mathcal{X}}_i$ is deterministic for every i with $\lambda_i^* = \widehat{\lambda}$.*

Proof. If $\widehat{\lambda} = 0$, this follows from Theorem 7.1(i).

Suppose now $\widehat{\lambda} > 0$, and consider the urn with an added dummy colour 0 as in Section 5. Then $\lambda_0^* = \widehat{\lambda}$ by (5.2). Furthermore, $\lambda_\nu = \widehat{\lambda} > 0$ and thus $a_\nu > 0$; hence $\nu \to 0$ and consequently ν is still the only leader k with $\lambda_k = \widehat{\lambda}$. In particular, $\mathsf{A}_0 = \{\nu\}$. Lemma 6.1 shows that $\mathcal{X}_i = c_{i\nu}\mathcal{X}_\nu$ for every $i \in \mathsf{Q}$ with $\lambda_i^* = \widehat{\lambda}$, and also that $\mathcal{X}_0 = c_{0\nu}\mathcal{X}_\nu$. Hence, the result follows from (5.12). ☐

Even when the limits $\widehat{\mathcal{X}}_i$ are not deterministic, there are generally linear dependencies between them just as for $\mathcal{X}_i$. In fact, in the main case $\widehat{\lambda} > 0$, we have the following analogue of Lemmas 6.1 and 6.7.

Lemma 7.3. *If $\widehat{\lambda} > 0$, then for every $i \in \mathsf{Q}$*

$$\widehat{\mathcal{X}}_i = \sum_{k \in \mathsf{A}_i} \widehat{c}_{ik}\widehat{\mathcal{X}}_k, \tag{7.1}$$

where

$$\widehat{c}_{ik} = \widehat{\lambda}^{-\kappa_i} c_{ik} > 0. \tag{7.2}$$

Moreover, Lemma 6.7 holds also for $\widehat{c}_{ik}$, with the only difference that (6.20) *is replaced by*

$$\widehat{c}_{i\nu} = \sum_{j \in \mathsf{D}_\nu^{\kappa-1}} \frac{a_j r_{ji}}{\kappa\widehat{\lambda}} \widehat{c}_{j\nu}. \tag{7.3}$$

Proof. The expansion (7.1)–(7.2) follows from (6.6) and (5.12), since $k \in \mathsf{A}_i$ implies $\lambda_k^* = \lambda_i^*$ and $\kappa_k = 0$. The final claim follows by (7.2) and Lemma 6.7. ☐

We leave the corresponding result for the less interesting case $\widehat{\lambda} = 0$ to the reader.

8. Urns with subtractions

We have so far assumed (A5): $\xi_{ij} \geqslant 0$ for all $i, j \in \mathsf{Q}$. In many applications, there are urns with subtractions, where we allow $\xi_{ij} < 0$. In the present paper, we are only interested in cases where the urn allows an infinite number of drawings according to the rules in Section 1; in other words we want (1.1) to make sense as probabilities for all n, and thus we require that the urn is such that $X_{ni} \geqslant 0$ for all n, and also $\sum_i a_i X_{ni} \neq 0$. Such urns are often called *tenable*. (See e.g. [38] for a discussion and examples.) A standard way to ensure that $X_{ni} \geqslant 0$ without assuming $\xi_{ii} \geqslant 0$ is to assume that X_{0i} and every ξ_{ji} always is an integer, so that X_{ni} is an integer, and that $\xi_{ii} \geqslant -1$ while $\xi_{ji} \geqslant 0$ for $j \neq i$. Typically, this is done for all i. (Thus $\mathbf{X}_n \in \mathbb{Z}^q_{\geqslant 0}$; we may then regard the urn process as drawings without replacement followed by adding $\xi_{ij} + \delta_{ij} \geqslant 0$ balls of each colour j.) However, we will be more flexible and allow a combination with this assumption for some colours i, and $\xi_{ji} \geqslant 0$ (as in earlier sections) for the others. We therefore assume:

(A5′) For each colour $i \in \mathsf{Q}$, we have either (or both)
(a) $\xi_{ji} \geqslant 0$ a.s. for all $j \in \mathsf{Q}$, or
(b) $\xi_{ii} \in \mathbb{Z}_{\geqslant 0} \cup \{-1\}$ a.s. and $\xi_{ji} \in \mathbb{Z}_{\geqslant 0}$ a.s. for all $j \neq i$, and $X_i(0) \in \mathbb{Z}_{\geqslant 0}$.

Note that (A5) is (A5′)(a) for every $i \in \mathsf{Q}$. Note also that (A5′) implies $\xi_{ji} \geqslant 0$ a.s. for all $i, j \in \mathsf{Q}$ with $i \neq j$. We let

$$\mathsf{Q}^- := \bigl\{i \in \mathsf{Q} : \mathbb{P}(\xi_{ii} < 0) > 0\bigr\} = \bigl\{i \in \mathsf{Q} : \mathbb{P}(\xi_{ii} = -1) > 0\bigr\} \tag{8.1}$$

denote the set of colours where (A5′)(b) but not (A5′)(a) applies. Note that (A2) implies that $a_i > 0$ for every $i \in \mathsf{Q}^-$.

Remark 8.1. More generally, we may assume that, for a given i, ξ_{ii} may take some negative value $-b$, provided it does not take any other negative value, and $X_i(0)$ and all ξ_{ji} ($j \in \mathsf{Q}$) a.s. are multiples of b. This case is easily reduced to the case $b = 1$ in (A5′) by dividing all X_{ni}, $X_i(t)$, and ξ_{ji} by b and multiplying the activity a_i by b. (See Example 14.8 for an example where we, however, use a slightly different alternative.) For convenience, and thus without real loss of generality, we will assume (A5′), where the only allowed negative replacement is -1. △

Remark 8.2. If $i \in \mathsf{Q}^- \cap \mathsf{Q}_{\min}$, then $X_i(t)$ is not influenced by the other colours, and (A5′)(b) implies that $X_i(t)$ is integer-valued and a classical continuous-time Markov branching process of the type studied in e.g. [4, Chapter III]. (As noted above, we have $a_i > 0$ by (A2).) Recall that [4, Section III.7] in the subcritical and critical cases $\lambda_i \leqslant 0$, this process a.s. dies out, i.e., $X_i(t) = 0$ for all sufficiently large t, while in the supercritical case $\lambda_i > 0$, the process survives for ever with positive probability (assuming $\mathbb{E}\,\xi_{ii} \log \xi_{ii} < \infty$, as we do); this probability is strictly less than 1, since the process always may die out when $\mathbb{P}(\xi_{ii} = -1) > 0$. △

We continue to use the definitions above, in particular (2.7)–(2.13). Note that we now may have $r_{ii} < 0$, and thus $\lambda_i < 0$; we may also have $\lambda_i = r_{ii} = 0$ without $\xi_{ii} = 0$ a.s.

We will see that the results in the previous sections are valid with minor changes also if we replace (A5) by (A5′), at least under some further technical assumptions (A6)–(A8) given below. However, in order to get more general results, these will be assumed only when needed. Hence, we assume throughout this section (A1)–(A4) and (A5′), with the following further assumptions added explicitly when needed.

(A6) $\sum_{i \in \mathsf{Q}} a_i X_i(t) > 0$ a.s. for all $t \geqslant 0$.
(A7) If $i \in \mathsf{Q}^-$, then $\lambda_i^* > 0$.
(A8) If $i \in \mathsf{Q}$ is a minimal colour, then $\xi_{ii} \geqslant 0$ a.s. (i.e., $i \notin \mathsf{Q}^-$).

In other words, (A7) says that if $i \in \mathsf{Q}^-$, then either $\lambda_i > 0$ or there exists $j \in \mathsf{Q}$ with $j \prec i$ and $\lambda_j > 0$ (or both). In particular, if i is a minimal colour, then either $\lambda_i > 0$ or $\xi_{ii} \geqslant 0$ a.s.

Note that (A8) implies that $X_i(t) \geqslant X_i(0)$ when i is minimal; since (A1)–(A3) imply that there exists a minimal i with $a_i > 0$ and $X_i(0) > 0$, (A8) together with (A1)–(A3) imply (A6).

8.1. **Some motivation for (A6)–(A8).** The assumption (A5′) implies that the continuous-time urn $\mathbf{X}(t) = (X_i(t))_{i \in \mathsf{Q}}$ is well-defined for $t \in [0, \infty)$, with $X_i(t) \geqslant 0$ for all $i \in \mathsf{Q}$. However, the urn might possibly reach a state where $X_i(t) = 0$ for all i with $a_i > 0$, and thus $\sum_i a_i X_i(t) = 0$; such a state is absorbing and no more draws will be made, and then the total number of draws is finite and the discrete-time urn is not well defined for all n. For results on the discrete-time urn, we therefore further assume (A6). Since $\sum_i a_i X_i(t)$ is constant between the jumps, (A6) implies that waiting times $\widehat{T}_{n+1} - \widehat{T}_n$ are finite a.s., and thus there is a.s. an infinite number of draws; hence $\widehat{T}_n < \infty$ for every n, and the discrete-time urn $(\mathbf{X}_n)_1^\infty$ is well defined by (2.19).

Next, we note that a colour i with $\lambda_i^* < 0$ will die out:

Lemma 8.3. *If $\lambda_i^* < 0$, then a.s. $X_i(t) = 0$ for all large t, and thus $X_{ni} = 0$ for all large n.*

Proof. Since every $j \to i$ has $\lambda_j^* \leqslant \lambda_i^* < 0$, we can use induction and assume that the statement holds for every $j \to i$. Then there exists a.s. a (random) time T such that no colour j with $j \to i$ exists for times $t \geqslant T$. Hence, for $t \geqslant T$, $X_i(t)$ evolves as a single colour urn with only colour i; we have $\lambda_i \leqslant \lambda_i^* < 0$, and thus, as in Remark 8.2, this process is a subcritical Markov branching process, which dies out a.s. □

Hence, if $\lambda_i^* < 0$, we can wait until colour i and all its ancestors have disappeared, and we may then regard the urn as restarted at that time; then we have an urn with fewer colours. Consequently, we may without loss of generality assume $\lambda_i^* \geqslant 0$ for every colour i. Note that it follows from the definition (2.8) that

$$\lambda_i^* \geqslant 0 \text{ for every colour } i \iff \lambda_j \geqslant 0 \text{ for every minimal colour } j. \tag{8.2}$$

Moreover, also the case $\lambda_i^* = 0$ may be problematic when $i \in \mathsf{Q}^-$, and we will actually make a stronger assumption than (8.2). There are two reasons.

First, suppose that $i \in \mathsf{Q}^-$ is a minimal colour (and thus $\lambda_i^* = \lambda_i$). Then, as said in Remark 8.2, $X_i(t)$ is a Markov branching process which dies out also in the critical case $\lambda_i = \lambda_i^* = 0$, so as above we may in this case wait until colour i has disappeared and consider an urn with fewer colours.

Secondly, and more importantly, if $i \in \mathsf{Q}^-$ is not minimal and $\lambda_i^* = 0$, there are examples where (after normalization) $X_i(t)$ converges in distribution but *not* a.s., and similarly for X_{ni}; see Examples 14.14 and 14.15. (In Example 14.15, X_{ni} does not even converge in distribution.)

We therefore exclude these cases and assume (A7). Note that if $i \notin \mathsf{Q}^-$, then $\lambda_i^* \geqslant \lambda_i \geqslant 0$. Hence, (A7) implies that $\lambda_i^* \geqslant 0$ for every i, and thus both conditions in (8.2) hold. (So we do not have to assume this explicitly.)

Even with these assumptions, one complication remains. If $i \in \mathsf{Q}^-$ is a minimal colour, then, as noted in Remark 8.2, $X_i(t)$ is a branching process, and even in the supercritical case $\lambda_i > 0$, it dies out with positive probability. We will accept this complication, but note that it may be eliminated by the additional assumption (A8), which we only assume when needed.

8.2. **Main results for urns with subtractions.** With the assumptions above, our main theorems for discrete and continuous time still hold:

Theorem 8.4. *Let $(X_i(t))_{i \in \mathsf{Q}}$ be a discrete-time triangular Pólya urn satisfying the conditions* (A1)–(A4), (A5′), *and* (A7)–(A8)*. Then the conclusions of Theorem 1.8 hold.*

We do not explicitly assume (A6) in Theorem 8.4, but it is implicit since it follows from (A8) and the other assumptions, as noted above.

Without (A8), we get a more complicated partial result (where $\mathcal{X}_0$ is as in Section 5); it is particularly useful in cases where $\mathcal{X}_0 > 0$ a.s. for some other reason, for example by Lemma 10.9 below in the balanced case. (See Example 14.8.)

Theorem 8.5. *Let $(X_i(t))_{i\in\mathsf{Q}}$ be a discrete-time triangular Pólya urn satisfying the conditions* (A1)–(A4), (A5′)*, and* (A6)–(A7)*. Then the conclusions of Theorem 1.8 hold a.s. on the event $\{\mathcal{X}_0 > 0\}$ (which has positive probability), except that $\widehat{\mathcal{X}}_i = 0$ is possible, with $0 \leqslant \mathbb{P}(\widehat{\mathcal{X}}_i = 0 \mid \mathcal{X}_0 > 0) < 1$.*

Theorem 8.6. *Let $(X_i(t))_{i\in\mathsf{Q}}$ be a continuous-time triangular Pólya urn satisfying the conditions* (A1)–(A4), (A5′)*, and* (A7)*. Then the conclusions of Theorem 4.1 hold, except that $\mathcal{X}_i = 0$ is possible, with $0 \leqslant \mathbb{P}(\mathcal{X}_i = 0 \mid \mathcal{X}_0 > 0) < 1$; moreover, we have $\mathbb{P}(\mathcal{X}_i > 0\ \forall i \in \mathsf{Q}) > 0$.*

If furthermore (A8) *holds, then the all conclusions of Theorem 4.1 hold.*

Remark 8.7. In Theorems 8.5 and 8.6, if the limit $\widehat{\mathcal{X}}_i = 0$ or $\mathcal{X}_i = 0$, then the result (1.3), (1.4) or (4.1) does not give the correct growth of X_{ni} or $X_i(t)$. In such cases, it might be possible to find more precise results using our methods on suitable subsets of the colours; it seems that this can be done very generally, at least on a case-by-case basis. We give one example in Example 14.13 but do not attempt to state any general theorem. △

Remark 8.8. Also the results in Section 6–7 still hold (with the same proofs) under the assumptions (A1)–(A4), (A5′), and (A7)–(A8), as the reader might verify.

If we do not assume (A8) and assume only (A1)–(A4), (A5′), and (A6)–(A7), then the results on absolute continuity do not hold as stated, since typically the limits may be 0 with positive probability. (We conjecture that these results might hold conditioned on the variables being non-zero, but we have not pursued this.) The remaining results in Section 6–7 still hold. △

Note that (A5) implies (A5′), (A7), and (A8) (with $\mathsf{Q}^- = \varnothing$), and thus (A1)–(A5) imply also (A6); hence these theorems (strictly) extend Theorems 1.8 and 4.1.

We prove Theorems 8.4–8.6 in the remainder of this section. As in Section 3, we first study a single colour, and as there we split the discussion into several subsections.

First note that we used that $X_j(t)$ is increasing in Lemma 3.1 and its proof. This is no longer necessarily true, but we may replace $X_j(t)$ by $X_j^*(t)$ with no further consequences. Hence, Lemmas 3.1 and 3.2 hold with the assumption modified to $\mathbb{E}\, X_j^*(t) < \infty$.

8.3. **A colour not influenced by others.** Consider the situation in Section 3.1, where $\xi_{ji} = 0$ for $j \neq i$, i.e., $i \in \mathsf{Q}_{\min}$. If $i \in \mathsf{Q}^-$, then $X_i(t)$ is, as noted in Remark 8.2, a Markov branching process.

Lemma 3.3 extends to this case, with some modifications.

Lemma 8.9. *Suppose that* (A1)–(A4) *and* (A5′) *hold and that $i \in \mathsf{Q}_{\min}$.*

(a) *If $i \notin \mathsf{Q}^-$ (i.e., $\xi_{ii} \geqslant 0$ a.s.), then Lemma 3.3 still holds.*

(b) *If $i \in \mathsf{Q}^-$, then Lemma 3.3 holds with the following modifications:*

  (i) *If $\lambda_i > 0$, the only difference is that $\mathcal{X}_i = 0$ is possible with positive probability. If $x_0 > 0$, then $0 < \mathbb{P}(\mathcal{X}_i = 0) < 1$.*

  (ii) *If $\lambda_i = 0$, then $X_i(t)$ is a martingale with*

$$X_i(t) \xrightarrow{\text{a.s.}} \mathcal{X}_i = 0, \qquad \text{as } t \to \infty, \tag{8.3}$$

$$\mathbb{E}\left[X_i(t)\right] = x_0, \tag{8.4}$$

$$\operatorname{Var}\left[X_i(t)\right] = Cx_0 t, \tag{8.5}$$

$$\mathbb{E}\left(\sup_{0\leqslant t\leqslant u} X_i(t)\right)^2 \leqslant Cx_0^2 + Cx_0 u, \qquad \textit{for every } u < \infty. \tag{8.6}$$

Furthermore, for every $\delta > 0$,

$$\mathbb{E}\left(\sup_{t\geqslant 0}\{e^{-\delta t}X_i(t)\}\right)^2 < \infty. \tag{8.7}$$

(iii) *If* $\lambda_i < 0$, *then*

$$X_i(t) \xrightarrow{\text{a.s.}} \mathcal{X}_i = 0, \qquad \textit{as } t \to \infty, \tag{8.8}$$

$$\operatorname{Var}\left[X_i(t)\right] \leqslant Cx_0 e^{\lambda_i t}, \tag{8.9}$$

$$\mathbb{E}\left(\sup_{0\leqslant t<\infty} X_i(t)\right)^2 \leqslant Cx_0^2. \tag{8.10}$$

Proof. (a): We thus assume (A5) = (A5′)(a) for the colour i, and then the proof of Lemma 3.3 still holds. (Recall that for simplicity we had (A1)–(A5) as standing assumptions in Section 3, including (A5) for all colours; however, as remarked before Lemma 3.3, this is needed only for the colour i in that subsection, and in particular for Lemma 3.3 and its proof.)

(b): Now assume (A5′)(b). By (A2), we also have $a_i > 0$.

We argue as in the proof of Lemma 3.3, with the following differences. Note that x_0 is assumed to be an integer, and the case $x_0 = 0$ is trivial; we thus may assume $x_0 \geqslant 1$. It is still true that $M(t) := e^{-\lambda_i t}X_i(t)$ is a martingale, and (3.20) holds if $\lambda_i \neq 0$.

Now, however, as said in Remark 8.2, with positive probability, $X_i(t) = 0$ for all large t; moreover, in the critical and subcritical cases (ii) and (iii), this happens a.s.

We study the three cases separately:

(i): The proof of Lemma 3.3 still holds, except for the final part yielding (3.25).

By Remark 8.2, with positive probability $X_i(t)$ dies out and thus $\mathcal{X}_i = 0$. Moreover, by (3.14), $\mathbb{E}\,\mathcal{X}_i = x_0 > 0$. Consequently, $0 < \mathbb{P}(\mathcal{X}_i = 0) < 1$.

(ii): By Remark 8.2, a.s. $X_i(t) = 0$ for all large t, which gives (8.3).

For $\lambda_i = 0$, (3.12) says that $X_i(t) = M(t)$ is a martingale, which implies (8.4). Furthermore, (3.20) in this case yields

$$\mathbb{E}\,X_i(t)^2 = \mathbb{E}\,M(t)^2 = \mathbb{E}\,[M,M]_t = x_0^2 + a_i\beta x_0 t, \tag{8.11}$$

which gives (8.5) and, together with Doob's inequality, (8.6). To prove (8.7), we note that (8.6) implies

$$\mathbb{E}\left(\sup_{t\geqslant 0}\{e^{-\delta t}X_i(t)\}\right)^2 \leqslant \mathbb{E}\sum_{n=0}^{\infty} e^{-2\delta n}\left(\sup_{n\leqslant t\leqslant n+1} X_i(t)\right)^2 \leqslant \mathbb{E}\sum_{n=0}^{\infty} e^{-2\delta n}O(n+1) < \infty. \tag{8.12}$$

(iii): As for (ii), Remark 8.2 shows that a.s. $X_i(t) = 0$ for all large t, and thus (8.8) holds. Since $\lambda_i < 0$, we obtain from (3.20)

$$\mathbb{E}\,M(t)^2 = \mathbb{E}\,[M,M]_t \leqslant x_0^2 + Cx_0 e^{-\lambda_i t} \tag{8.13}$$

and thus

$$\operatorname{Var}\left[X_i(t)\right] = e^{2\lambda_i t}\operatorname{Var}\left[M(t)\right] \leqslant Cx_0 e^{\lambda_i t}, \tag{8.14}$$

showing (8.9). Furthermore, (8.13) and Doob's inequality yield, for any $m \geqslant 1$,

$$\mathbb{E}\big[\sup_{m-1\leqslant t\leqslant m} X_i(t)^2\big] \leqslant e^{2\lambda_i(m-1)}\,\mathbb{E}\big[\sup_{m-1\leqslant t\leqslant m} M(t)^2\big] \leqslant Ce^{2\lambda_i m}\,\mathbb{E}\big[M(m)^2\big] \leqslant Ce^{\lambda_i m}x_0^2. \tag{8.15}$$

Since $\lambda_i < 0$, the sum over all $m \geqslant 1$ is $\leqslant Cx_0^2$, which implies (8.10). ☐

Remark 8.10. In fact, it follows from [4, Theorem III.7.2] (using (A4)) that if $i \in \mathsf{Q}_{\min} \cap \mathsf{Q}^-$, then $\mathcal{X}_i = 0$ occurs a.s. exactly when $X_i(t) = 0$ for large t so that the branching process dies out in finite time. Thus, $\mathbb{P}(\mathcal{X}_i = 0)$ equals the probability that the continuous-time branching process $X_i(t)$ dies out. Considering only the times that a ball of colour i is drawn, we obtain an embedded random walk with i.i.d. increments distributed as ξ_{ii}; thus $\mathbb{P}(\mathcal{X}_i = 0)$ also equals the probability that this random walk hits 0. (It is easy to see that this also equals the probability of extinction for a Galton–Watson process with offspring distribution $1 + \xi_{ii}$ started with x_0 individuals.) △

8.4. **A colour only produced by one other colour.** Consider the situation in Section 3.2, where there is a single colour j such that $j \to i$.

We used that $X_j(t)$ is increasing in (3.56); this is no longer necessarily true, but we may replace $X_j(t)$ by $X_j^*(t)$ there with no further consequences.

Lemma 3.5 holds with the following minor modifications. We assume $\lambda_j^* \geqslant 0$, since otherwise j eventually becomes extinct by Lemma 8.3, and after that $X_i(t)$ evolves as in Lemma 8.9. (We are not really interested in this case, as discussed earlier.) Note that we also exclude the case $\lambda_i^* = \lambda_j^* = 0$. (For good reasons, see Examples 14.14 and 14.15).

Lemma 8.11. *Suppose that* (A1)–(A4) *and* (A5′) *hold. Let* $i \in \mathsf{Q}$, *and suppose that there is exactly one colour* $j \in \mathsf{Q}$ *such that* $j \to i$, *and that* $X_i(0) = 0$. *Suppose also that one of the following holds:*

(a) $i \notin \mathsf{Q}^-$ *(i.e.,* $\xi_{ii} \geqslant 0$ *a.s.), and* $\lambda_j^* \geqslant 0$.
(b) $i \in \mathsf{Q}^-$, $\lambda_i^* > 0$, *and* $\lambda_j^* \geqslant 0$.

Then Lemma 3.5 still holds, i.e., if (3.31)–(3.32) *hold, then we have the conclusions* (3.33)–(3.35) *and* $\mathcal{X}_j > 0 \implies \mathcal{X}_i > 0$ *a.s.*

Proof. (a): Recall again that we had (A1)–(A5) as standing assumptions in Section 3, including (A5) for all colours; however, it is easily verified that the proof of Lemma 3.5 does not use this for other colours than i, except in (3.56), where we now replace $X_j(t)$ by $X_j^*(t)$ as discussed above, and to see that $\lambda_j^* \geqslant 0$. Hence, in the present setting where we assume (A1)–(A4) and (A5′), and also explicitly assume $\lambda_j^* \geqslant 0$, if $\xi_{ii} \geqslant 0$ a.s., then the proof of Lemma 3.5 still holds.

(b): We now assume (A5′)(b) for i. Most of the proof of Lemma 3.5 remains the same. The main difference in the proof comes in Steps 4–5, where we used Lemma 3.3 in (3.65), but now instead use Lemma 8.9(b). We consider three cases:

(1) $\lambda_i > 0$: Then (3.65) still holds by Lemma 8.9, and thus all estimates in Steps 4–5 hold.
(2) $\lambda_i = 0$: Then (3.65) is replaced by, using (8.5),

$$\mathbb{E}\big[[Z_1, Z_1]_t \mid (T_k, \eta_k)_1^\infty\big] = \sum_{k=1}^\infty \mathbf{1}_{\{T_k\leqslant t\}}\operatorname{Var}\big[Y_k(t - T_k) \mid T_k, \eta_k\big]$$

$$= \sum_{k=1}^{\infty} \mathbf{1}_{\{T_k \leqslant t\}} C(t - T_k)\eta_k. \tag{8.16}$$

Thus

$$\mathbb{E}\,[Z_1, Z_1]_t = C\,\mathbb{E} \sum_{k=1}^{\infty} \mathbf{1}_{\{T_k \leqslant t\}} (t - T_k) \leqslant Ct\,\mathbb{E}\, N(t). \tag{8.17}$$

Using Lemma 3.2 (or Lemma 3.1), it follows that (3.68) for $\ell = 1$ is replaced by (with $\lambda_i = 0$)

$$\mathbb{E}\,[Z_1, Z_1]_t \leqslant Ct\,\mathbb{E} \int_0^t X_j(s)\,\mathrm{d}s \leqslant C\,\mathbb{E}\,\widetilde{X}_j^{**} t \int_0^t (s+1)^{\kappa_j} e^{(\lambda_j^* - 2\lambda_i)s}\,\mathrm{d}s, \tag{8.18}$$

with an extra factor t. We assume $0 < \lambda_i^* = \lambda_i \vee \lambda_j^*$ and $\lambda_i = 0$, and thus $\lambda_j^* > 0 = \lambda_i$, so we are in Case (ii$'$) of Step 5; (3.71) now holds with an insignificant extra factor n and thus (3.72)–(3.73) still hold and the conclusions of Step 5 remain the same.

(3) $\lambda_i < 0$: Then we obtain instead of (3.65), now using (8.9),

$$\begin{aligned}\mathbb{E}\big[[Z_1, Z_1]_t \mid (T_k, \eta_k)_1^{\infty}\big] &\leqslant \sum_{k=1}^{\infty} \mathbf{1}_{\{T_k \leqslant t\}} e^{-2\lambda_i T_k} C e^{-\lambda_i (t - T_k)} \eta_k \\ &= C \sum_{k=1}^{\infty} \mathbf{1}_{\{T_k \leqslant t\}} e^{-\lambda_i t - \lambda_i T_k} \eta_k. \end{aligned} \tag{8.19}$$

Thus, by Lemma 3.2(ii),

$$\mathbb{E}\,[Z_1, Z_1]_t \leqslant C\,\mathbb{E} \sum_{k=1}^{\infty} \mathbf{1}_{\{T_k \leqslant t\}} e^{-\lambda_i t - \lambda_i T_k} \eta_k = C e^{-\lambda_i t}\,\mathbb{E} \int_0^t e^{-\lambda_i s} X_j(s)\,\mathrm{d}s. \tag{8.20}$$

It follows that (3.68) for $\ell = 1$ is replaced by

$$\mathbb{E}\,[Z_1, Z_1]_t \leqslant C\,\mathbb{E}\,\widetilde{X}_j^{**} e^{-\lambda_i t} \int_0^t (s+1)^{\kappa_j} e^{(\lambda_j^* - \lambda_i)s}\,\mathrm{d}s. \tag{8.21}$$

However, for $t \leqslant 1$, (8.21) is trivially equivalent to (3.68), and for $t \geqslant 1$, (8.21) implies, using $\lambda_j^* - \lambda_i \geqslant -\lambda_i > 0$,

$$\begin{aligned}\mathbb{E}\,[Z_1, Z_1]_t &\leqslant C\,\mathbb{E}\,\widetilde{X}_j^{**} (t+1)^{\kappa_j} e^{(\lambda_j^* - 2\lambda_i)t} \\ &\leqslant C\,\mathbb{E}\,\widetilde{X}_j^{**} \int_{t-1}^t (s+1)^{\kappa_j} e^{(\lambda_j^* - 2\lambda_i)s}\,\mathrm{d}s. \end{aligned} \tag{8.22}$$

Hence (3.68) holds for all $\ell \leqslant 3$ also in the present setting. By our assumption $\lambda_i^* = \lambda_i \vee \lambda_j^* > 0$ we have $\lambda_j^* > 0$, so we are again in Case (ii$'$) of Step 5; (3.71)–(3.73) still hold and the conclusions of Step 5 remain the same.

In all cases, the conclusions (3.33)–(3.35) of Step 6 follow as for Lemma 3.5.

Finally, as for Lemma 3.5, if $\lambda_i \leqslant \lambda_j^*$, then (3.35) shows that $\mathcal{X}_j > 0 \implies \mathcal{X}_i > 0$. If $\lambda_i > \lambda_j^*$ and $\mathcal{X}_j > 0$, then we use again (3.78), where still a.s. all T_k are finite. Let $\mathcal{Y}_k := \lim_{t\to\infty} e^{-\lambda_i t} Y_k(t)$. In contrast to Lemma 3.5, $\mathcal{Y}_k = 0$ is now possible also if $\eta_k > 0$, but $0 < \mathbb{P}(\mathcal{Y}_k = 0) < 1$ by Lemma 8.9. We now let K be the smallest k such that $\mathcal{Y}_k > 0$, and conclude as in Section 3 that $\mathcal{X}_i > 0$. ☐

Remark 8.12. In Lemma 8.11(b) we have excluded the case $\lambda_i = \lambda_j^* = 0$, since then $\lambda_i^* = 0$. In this case we are in Case (iii′) of Step 5 of the proof of Lemma 3.5. As mentioned above, there is now an extra factor t in (3.68), and therefore we obtain instead of (3.75)

$$\mathbb{E}\,\widetilde{Z}_\ell^{\ddagger}(n)^2 \leqslant \overline{C} 2^{-\kappa_j n}. \tag{8.23}$$

Hence, if furthermore $\kappa_j \geqslant 1$, then (3.76) still holds, and the rest of the proof of Lemma 3.5 applies. Consequently, Lemma 8.11(b) holds also in the case $\lambda_i = \lambda_j^* = 0$ and $\kappa_i \geqslant 1$.

Example 14.14 gives an example with $\lambda_i = \lambda_j^* = 0 = \kappa_i$ where the conclusion of Lemma 8.11 does not hold. For simplicity we have in Lemma 8.11(b) excluded all cases with $\lambda_i = \lambda_j^* = 0$ (and thus $\lambda_i^* = 0$).

Similarly, Example 14.15 gives an example with $\lambda_i < 0 = \lambda_j^* = \kappa_i$ where the conclusion of Lemma 8.11 does not hold. We have excluded all cases with $\lambda_i < \lambda_j^* = 0$; it seems likely that Lemma 8.11(b) might hold also in this case if $\kappa_i \geqslant 1$, but we have not investigated this further. △

We need also a coarse estimate in the case excluded in Lemma 8.11(b).

Lemma 8.13. *Suppose that* (A1)–(A4) *and* (A5′) *hold and that* $i \in \mathsf{Q}$ *is such that* $\mathbb{P}(\xi_{ii} = -1) > 0$. *Assume that there is exactly one colour* $j \in \mathsf{Q}$ *such that* $j \to i$, *and that* $X_i(0) = 0$. *Assume also that* $\lambda_i^* = \lambda_j^* = 0$.

Assume further that (3.31)–(3.32) *hold. Then, for every* $\delta > 0$,

$$e^{-\delta t} X_i(t) \xrightarrow{\text{a.s.}} 0, \tag{8.24}$$

and

$$\Bigl\|\sup_{t\geqslant 0}\bigl\{e^{-\delta t} X_i(t)\bigr\}\Bigr\|_2 < \infty. \tag{8.25}$$

Proof. We have $\lambda_i \leqslant \lambda_i^* = 0$ and, by (A2), $a_i > 0$. Thus $r_{ii} = \lambda_i / a_i \leqslant 0$. Furthermore, $\xi_{ii} \geqslant -1$ a.s., and thus $r_{ii} = \mathbb{E}\,\xi_{ii} \geqslant -1$. Choose $p \in (0,1)$ such that $0 < r_{ii} + 2p < \delta / a_i$.

Modify the urn by increasing ξ_{ii} to $\breve{\xi}_{ii} := \xi_{ii} + 2\zeta$, where $\zeta \in \mathrm{Be}(p)$ is independent of ξ_{ii}; all other $\xi_{k\ell}$ and initial conditions $X_k(0)$ remain the same. We denote quantities for the new urn by adding $\breve{\ }$. The modification does not affect any ancestors of i; thus $\breve{X}_j(t) = X_j(t)$ and $\breve{\lambda}_j^* = \lambda_j^* = 0$. On the other hand, by our choice of p,

$$\breve{\lambda}_i = a_i\,\mathbb{E}\,\breve{\xi}_{ii} = a_i(\mathbb{E}\,\xi_{ii} + 2\,\mathbb{E}\,\zeta) = a_i(r_{ii} + 2p) \in (0, \delta). \tag{8.26}$$

Consequently, $\breve{\lambda}_i^* = \breve{\lambda}_i \vee \breve{\lambda}_j^* = \breve{\lambda}_i > 0$ and thus Lemma 8.11(b) applies to the modified urn. (The other conditions of Lemma 8.11(b) obviously hold.)

The modification adds extra balls of colour i, and these may get descendants of colour i and they may disappear again, but we may separate these extra balls of colour i from the original ones, and thus couple the two versions such that $\breve{X}_i(t) \geqslant X_i(t)$ for all $t \geqslant 1$. Consequently, noting that $\breve{\lambda}_i^* = \breve{\lambda}_i$ and $\breve{\kappa}_i = 0$, Lemma 8.11 yields, from (3.33)–(3.34) and (3.29)–(3.30),

$$\limsup_{t\to\infty} e^{-\breve{\lambda}_i t} X_i(t) \leqslant \lim_{t\to\infty} e^{-\breve{\lambda}_i t} \breve{X}_i(t) = \breve{\mathcal{X}}_i < \infty \quad \text{a.s.}, \tag{8.27}$$

and

$$\Big\| \sup_{t\geqslant 0}\{e^{-\check\lambda_i t}X_i(t)\}\Big\|_2 \leqslant \Big\|\sup_{t\geqslant 0}\{e^{-\check\lambda_i t}\check X_i(t)\}\Big\|_2 < \infty. \tag{8.28}$$

The results follow, since $\check\lambda_i < \delta$ by (8.26). □

8.5. **The general case for a single colour.**

Lemma 8.14. *Suppose that* (A1)–(A4), (A5′), *and* (A7) *hold. Then Lemma 3.7 still holds, except for the last sentence.*

The last sentence remains valid if $i \notin \mathsf{Q}^-$, and also if $i \notin \mathsf{Q}_{\min}$. (In particular, it is valid if (A8) *holds.) In the remaining case, when $i \in \mathsf{Q}^-$ and is minimal, then* $0 < \mathbb{P}(\mathcal{X}_i > 0) < 1$.

Proof. Recall that the assumption (A7) implies that $\lambda_j^* \geqslant 0$ for every $j \in \mathsf{Q}$.

We follow the proof of Lemma 3.7 and split the colour i into subcolours i_0 and i_j, $j \in \mathsf{P}_i$ (also if i is minimal so $\mathsf{P}_i = \varnothing$); then (3.84) holds, and also (3.85)–(3.89) and (3.91).

If $i \notin \mathsf{Q}^-$, then the proof is exactly as for Lemma 3.7, now using Lemmas 8.9(a) and 8.11(a) instead of Lemmas 3.3 and 3.5.

Assume thus in the sequel of the proof that instead $i \in \mathsf{Q}^-$. (Thus, (A5′)(b) holds.) Then by (A7), we have $\lambda_i^* > 0$. However, it is possible that $\lambda_{i_0} = \lambda_i \leqslant 0$ or that $\lambda_{i_j}^* = \lambda_i \vee \lambda_j^* = 0$ for some $j \in \mathsf{P}_i$.

For i_0, we now apply Lemma 8.9(b). If $\lambda_i > 0$, then Lemma 8.9(b)(i) shows that X_{i_0} can be treated as before in (3.89) and (3.91). If $\lambda_i \leqslant 0$, we replace, as we may, $\widetilde X_{i_0}(t)$ by $e^{-\lambda_i^* t}X_{i_0}(t)$ in the estimate (3.89), cf. (3.84) and (3.29). By (8.3) and (8.8), this term contributes 0 to the limit (3.82). Similarly, in (3.91) we replace $\widetilde X_{i_0}^{**}$ by $\sup_{t\geqslant 0}\{e^{-\lambda_i^* t}X_{i_0}(t)\}$, which gives a finite contribution to (3.83) by (8.7) and (8.10).

Similarly, for each i_j ($j \in \mathsf{P}_i$), we now apply Lemma 8.11(b) if $\lambda_{i_j}^* > 0$; otherwise, i.e., if $\lambda_i \leqslant 0$ and $\lambda_j^* = 0$, we apply Lemma 8.13 (with $\delta := \lambda_i^* > 0$). In the latter case we replace $\widetilde X_{i_j}(t)$ by $e^{-\lambda_i^* t}X_{i_j}(t)$ in the estimate (3.89) and use (8.24), and we replace $\widetilde X_{i_j}^{**}$ by $\sup_{t\geqslant 0}\{e^{-\lambda_i^* t}X_{i_j}(t)\}$ in (3.91) and use (8.25).

This proves, in all cases, that (3.82) and (3.83) hold. Moreover, for each $j \in \mathsf{P}_i \cup \{0\}$ such that we apply Lemma 8.9(b)(ii),(iii) or Lemma 8.13 to i_j, the contribution to the limit (3.82) is 0, and so is the contribution to the right-hand side of (3.90) since in these cases $\lambda_{i_j}^* \leqslant 0 < \lambda_i^*$ by (3.85)–(3.86). Hence, (3.90) still holds.

It remains to consider the final sentence in Lemma 3.7. First, if $\mathsf{P}_i \neq \varnothing$, then it follows as in the proof of Lemma 3.7 that if $\mathcal{X}_j > 0$ for every $j \in \mathsf{P}_i$, then $\mathcal{X}_i > 0$. In the remaining case, i is minimal. Note that then $X_i(0) > 0$ by (A3). If i is minimal and $i \notin \mathsf{Q}^-$, then a.s. $\mathcal{X}_i > 0$ by Lemma 3.3. On the other hand, if i is minimal and $i \in \mathsf{Q}^-$, then Lemma 8.9(b)(i) shows that $\mathbb{P}(\mathcal{X}_i = 0) \in (0, 1)$. □

8.6. **Proofs of Theorems 8.4–8.6.**

Proof of Theorem 8.6. As in Section 4, now using Lemma 8.14; we use again induction on the colour i and conclude that (3.82) and (3.83) hold for every $i \in \mathsf{Q}$, although now $\mathcal{X}_i = 0$ is possible.

The processes $X_i(t)$ for $i \in \mathsf{Q}_{\min}$ are independent, and thus so are their limits $\mathcal{X}_i$. Each is strictly positive with positive probability, by Lemma 8.9 or 8.14, and thus there is a positive probability that $\mathcal{X}_i > 0$ for every minimal i. Moreover, if (A8) holds, then this probability is 1. Finally, it follows by Lemma 8.14 by induction on the colour i that a.s. if $\mathcal{X}_i > 0$ for every minimal i, then $\mathcal{X}_i > 0$ for every i. □

Proof of Theorems 8.4 and 8.5. As in Section 5, now using Theorem 8.6 and considering only the event $\{\mathcal{X}_0 > 0\}$. Note that Theorem 8.6, applied to the extended urn, shows that with positive probability $\mathcal{X}_0 > 0$ and $\mathcal{X}_i > 0$ for all $i \in \mathsf{Q}$; furthermore, in Theorem 8.4, where we assume (A8), this holds a.s.

Note also that since we assume (A6), we have a.s. $\widehat{T}_n < \infty$ for every n, and thus the discrete-time process $\mathbf{X}_n$ is well-defined. □

9. Random vs non-random replacements

Consider a Pólya urn $\mathcal{U}$ with a random replacement matrix $(\xi_{ij})_{i,j\in\mathsf{Q}}$. For simplicity we study only the case when all $\xi_{ij} \geqslant 0$, and we thus assume (A1)–(A5). Consider also another urn $\mathcal{U}'$ with the same colours Q, the same initial vector $\mathbf{X}_0$, and the same activities a_i, but non-random replacements $r_{ij} = \mathbb{E}\,\xi_{ij}$. We thus replace the replacement matrix by its mean, and we may call $\mathcal{U}'$ the *mean urn* corresponding to $\mathcal{U}$. Both urns have the same r_{ij}, and thus the same colour graph, λ_i, λ_i^*, κ_i, $\widehat{\lambda}$, $\widehat{\kappa}$ and γ_i, see (2.3) and (2.7)–(2.13). The new urn $\mathcal{U}'$ also satisfies (A1)–(A5), and Theorems 1.8 and 4.1 show that we have the same qualitative asymptotic behaviour for both urns, with the same normalization factors. However, the limits $\widehat{\mathcal{X}}_i$ and $\mathcal{X}_i$ are in general *not* the same for the two urns $\mathcal{U}$ and $\mathcal{U}'$, and thus the asymptotic distributions may be be different; see Example 14.5.

We note that the constants c_{ik} in Section 6 by Lemma 6.7 are the same for the two urns. This yields one simple case. (We assume $\widehat{\lambda} > 0$ for simplicity, and leave a study of the case $\widehat{\lambda} = 0$ to the reader.)

Theorem 9.1. *Assume $\widehat{\lambda} > 0$. If i is a colour such that $\widehat{\mathcal{X}}_i$ is deterministic for one of the two urns, then $\widehat{\mathcal{X}}_i$ is the same constant for both urns.*

Proof. By Theorem 7.1, there are two cases: either $\lambda_i^* = 0$, or $\lambda_i^* = \widehat{\lambda}$.

If $\lambda_i^* = 0$, then (6.5) and Lemma 6.2 show that for every $j \in \mathsf{A}_i$, $\mathcal{X}_j = X_j(0)$ is a constant, the same for both urns. Hence, Lemmas 6.1 shows, since the constants c_{ik} are the same for both urns, that $\mathcal{X}_i$ is the same constant for both urns. Consequently, (5.12) shows that $\widehat{\mathcal{X}}_i = \widehat{\lambda}^{-\kappa_i}\mathcal{X}_i$ is the same constant for both urns.

Assume now $\lambda_i^* = \widehat{\lambda}$. If we add dummy balls to both urns as in Section 5, then, by Lemma 6.1 and (5.2), for any of the urns, both $\mathcal{X}_i$ and $\mathcal{X}_0$ are linear combinations $\sum_{j\in J} c_{ij}\mathcal{X}_j$ and $\sum_{j\in J} c_{0j}\mathcal{X}_j$, where J is the set of leaders with $\lambda_j^* = \widehat{\lambda}$. The joint distribution of $(\mathcal{X}_j)_{\in J}$ is absolutely continuous by Theorem 6.4, and since (5.12) yields $\widehat{\mathcal{X}}_i = \widehat{\lambda}^{-\gamma_i}\mathcal{X}_i/\mathcal{X}_0$, it follows that if $\widehat{\mathcal{X}}_i$ is deterministic in one of the two urns, then the vectors $(c_{ij})_{j\in J}$ and $(c_{0j})_{j\in J}$ are proportional. Since these vectors are the same for both urns, it follows that then $\widehat{\mathcal{X}}_i$ is the same constant in both urns. □

10. Balanced urns

Many papers on Pólya urns assume that urn is balanced (as defined below); while this is quite restrictive, it is partly justified by the fact that Pólya urns that appear in applications often are balanced. We have not needed this assumption in the preceding sections, but it will be used for some results in the following sections (in particular Section 12). The present section contains the definition and some simple results for later use.

In the standard case when all activities $a_i = 1$, a Pólya urn is said to be *balanced* if we add the same total number of balls at each draw; in other words, if the replacement matrix $(\xi_{ij})_{i,j \in \mathsf{Q}}$ satisfies $\sum_j \xi_{ij} = b$ a.s., for some $b \in \mathbb{R}$ and every $i \in \mathsf{Q}$. In general, with possibly different activities a_i, it is not the total number of balls but their total activity that is important, and we make the following general definition. (Definition 10.1 and Lemma 10.2 – Remark 10.6 below apply to all Pólya urns, triangular or not.)

Definition 10.1. A Pólya urn is *balanced* if there exists a constant $b \in \mathbb{R}$ (called the *balance*) such that for every $i \in \mathsf{Q}$ with $a_i > 0$, we have

$$\sum_{j \in \mathsf{Q}} a_j \xi_{ij} = b \qquad \text{a.s.} \tag{10.1}$$

With $\mathbf{a} := (a_j)_{j \in \mathsf{Q}}$, the vector of activities, (10.1) may be written $\mathbf{a} \cdot \boldsymbol{\xi}_i = b$ for all i with $a_i > 0$.

Suppose that the urn is balanced as in Definition 10.1. It follows from (1.2) (since colours i with $a_i = 0$ will never be drawn) that a.s.

$$\mathbf{a} \cdot \mathbf{X}_n = \sum_{j \in \mathsf{Q}} a_j X_{nj} = \mathbf{a} \cdot \mathbf{X}_0 + nb. \tag{10.2}$$

Hence, the denominator in (1.1) is deterministic. This has been used in several ways in many papers by different authors. (In particular, it makes it possible to use martingale methods for the discrete-time urn, similar to the continuous-time arguments in the present paper; see Remark 1.11 and e.g. [12].) Here we note some simple consequences.

Note first that if the balance $b < 0$, then (10.2) shows that the activity $\mathbf{a} \cdot \mathbf{X}_n$ becomes negative for large n; this is clearly impossible and shows that the discrete-time urn process must stop and cannot be continued for ever. In the present paper, we are not interested in this case, so we must have $b \geqslant 0$, and we thus assume this in the sequel (whether it is said explicitly or not). In this case, there are no problems.

Lemma 10.2. *A balanced urn with balance $b \geqslant 0$ which satisfies* (A1) *and* (A5) *or (more generally)* (A5′) *is well-defined for all discrete or continuous times, and thus satisfies* (A6)*.*

Proof. By (10.2), we have, for every n, a.s. $\mathbf{a} \cdot \mathbf{X}_n \geqslant \mathbf{a} \cdot \mathbf{X}_0 > 0$. Hence the discrete-time process $\mathbf{X}_n$ is well-defined, and also $\mathbf{a} \cdot \mathbf{X}(t) > 0$ for every t. ☐

We assume (A1) and (A5) or (A5′) below; thus (A6) holds.

Remark 10.3. The case $b = 0$ is trivial. First, if (A5) holds, so there are no subtractions, then the only possibility is $\xi_{ij} = 0$ a.s. for all $i, j \in \mathsf{Q}$; in other words, we draw from the urn but nothing happens. If we allow subtractions as in (A5′)(b), there may be a few initial draws that wipe out some colours, but nothing more will happen; furthermore, such urns violate (A7). △

Lemma 10.4. *If a Pólya urn is balanced, then the random processes $(\mathbf{X}_n)_{n=1}^{\infty}$ and $(\widehat{T}_n)_{n=1}^{\infty}$ are independent.*

Proof. Let $n \geqslant 0$. Stop the continuous-time process $\mathbf{X}(t)$ at $\widehat{T}_n$, i.e., at the nth draw. Conditionally on everything that has happened so far, the waiting time $\widehat{T}_{n+1} - \widehat{T}_n$ until the next draw is an exponential random variable with rate $\sum_i a_i X_{ni}$, which by (10.2) is the constant $\mathbf{a} \cdot \mathbf{X}_0 + nb$. Hence, $\widehat{T}_{n+1} - \widehat{T}_n$ is independent of $\mathbf{X}_n$, and thus also of the colour of the drawn ball, say i, and of $\mathbf{X}_{n+1} - \mathbf{X}_n = \boldsymbol{\xi}_i^{(n)}$, recall (1.2). Consequently, the processes $(\widehat{T}_n)_n$ and $(\mathbf{X}_n)_n$ evolve independently. ☐

Lemma 10.5. *If a Pólya urn is balanced, then the distribution of the random sequence* $(\widehat{T}_n)_{n=1}^{\infty}$ *depends only on the balance* b *and the initial activity* $\mathbf{a} \cdot \mathbf{X}_0$.

Proof. As in the proof of Lemma 10.4, $\widehat{T}_{n+1} - \widehat{T}_n$ is exponential distributed with rate $\mathbf{a} \cdot \mathbf{X}_0 + nb$, and these waiting times are independent. ☐

Remark 10.6. The distribution of $(\widehat{T}_n)_{n=1}^{\infty}$ is thus the same as for the single-colour urn with initial value $x_0 := \mathbf{a} \cdot \mathbf{X}_o$, activity $a_1 = 1$, and replacement matrix (b). For $b > 0$, this is the more or less trivial case $q = 1$ of Example 14.1, and it is well known that $(b\widehat{T}_n)_1^{\infty}$ are the jump times of a Yule process started with x_0/b individuals. △

In the remaining part of the section we consider, as in the rest of the paper, triangular urns.

Lemma 10.7. *Consider a triangular balanced urn with balance* $b \geqslant 0$, *and suppose that the urn satisfies* (A1)–(A4). *Then* $\widehat{\lambda} = b$. *Furthermore, if* $b > 0$, *then* $\widehat{\kappa} = 0$.

Proof. Let $\mathsf{Q}' := \{i \in \mathsf{Q} : a_i > 0\}$. Then $\mathsf{Q}' \neq \varnothing$ by (A1). There are three possibilities for a colour $i \in \mathsf{Q}$:

(i) If $i \notin \mathsf{Q}'$, then $a_i = 0$ and thus $\lambda_i = 0$.

(ii) If $i \in \mathsf{Q}'$ and i is maximal in Q' for the order $\prec$, i.e., $i \not\to j$ for every $j \in \mathsf{Q}'$, then $\xi_{ij} = 0$ a.s. when $j \in \mathsf{Q}' \setminus \{i\}$, and $a_j = 0$ for every $j \notin \mathsf{Q}'$; hence $a_j \xi_{ij} = 0$ a.s. for all colours $j \neq i$. Consequently, (10.1) yields

$$a_i \xi_{ii} = \sum_{j \in \mathsf{Q}} a_j \xi_{ij} = b \tag{10.3}$$

a.s., and thus, taking the expectation,

$$\lambda_i = a_i r_{ii} = a_i \operatorname{\mathbb{E}} \xi_{ii} = b. \tag{10.4}$$

(iii) If $i \in \mathsf{Q}'$ is not maximal in Q', then there exists $j \in \mathsf{Q}'$ with $i \to j$ and thus $r_{ij} > 0$. Taking expectations in (10.1) yields

$$\lambda_i = a_i r_{ii} = b - \sum_{j \neq i} a_j r_{ij} \leqslant b - a_j r_{ij} < b. \tag{10.5}$$

It follows that $\widehat{\lambda} := \max_i \lambda_i = b$. Furthermore, if $b > 0$, then the maximum is attained precisely in Case (ii), i.e., for i that are maximal in Q'. It follows that for every such i we have $\lambda_j < b$ for all $j \in \mathsf{P}_i$, and thus $\kappa_i = 0$. Hence, $\widehat{\kappa} = 0$. ☐

Remark 10.8. The proof shows also that if a balanced triangular urn has $b > 0$ and all activites $a_i = 1$, then $\xi_{ii} = r_{ii} = \lambda_i = b$ a.s. for every i that is maximal in Q', but $\operatorname{\mathbb{E}} \xi_{ii} = r_{ii} = \lambda_i < b$ for all other colours i. △

For urns with subtractions as in Section 8, there are further simplifications when the urn is balanced.

Lemma 10.9. *Consider a triangular balanced urn with balance $b \geqslant 0$, and suppose that the urn satisfies* (A1)–(A4), (A5′), *and* (A7). *Then, with notation as in Section 5, $\mathcal{X}_0 > 0$ a.s.*

Proof. The case $b = 0$ is trivial by Remark 10.3, so we may assume $b > 0$.

For convenience, denote the urn by $\mathcal{U}$. Lemma 10.5 and Remark 10.6 show that $\widehat{T}_n$ are (jointly) distributed as for the one-colour urn $\mathcal{U}_1$ with replacement matrix (b), activity $a = 1$, and the same initial total activity. Hence we may couple the urns such that they have the same $\widehat{T}_n$. (Or we may simply define the content of $\mathcal{U}_1$ as being the total activity in the urn $\mathcal{U}$.) The urn $\mathcal{U}_1$ obviously satisfies (A1)–(A5), and thus (5.5) applies to it, with $\mathcal{X}_0 > 0$ a.s. Moreover, (5.5) applies also to the original urn $\mathcal{U}$ by the same argument in Section 5 (as in the proof of Theorems 8.4 and 8.5 in Section 8.6); note that the two urns $\mathcal{U}$ and $\mathcal{U}_1$ have the same $\widehat{\lambda} = b$ and $\kappa_0 = 0$ by Lemma 10.7 and (5.3). Consequently, the two urns $\mathcal{U}$ and $\mathcal{U}_1$ have the same $\mathcal{X}_0$. ☐

In fact, in Lemma 10.9, $\mathcal{X}_0$ has a Gamma distribution by the proof and (14.2).

11. The drawn colours

We have so far studied $\mathbf{X}_n$ and $\mathbf{X}(t)$, the numbers of balls of each colour in the urn. It is also of interest to study the number of times each colour is drawn. (See Examples 14.7–14.9 for an application.) For $i \in \mathsf{Q}$, we denote the number of times that a ball of colour i is drawn up to time n in the discrete-time urn by N_{ni}, and the number of times up to time t in the continuous-time urn by $N_i(t)$; thus

$$N_{ni} = N_i(\widehat{T}_n), \qquad i \in \mathsf{Q},\ n \geqslant 0. \tag{11.1}$$

We state first a continuous-time and then a discrete-time result; both are similar to the results for $\mathbf{X}(t)$ and $\mathbf{X}_n$ earlier, but note that exponents change when $\lambda_i^* = 0$. (Proofs are given later in this section.)

Theorem 11.1. *Let $(X_i(t))_{i\in\mathsf{Q}}$ be a continuous-time triangular Pólya urn satisfying either* (A1)–(A5), *or (more generally)* (A1)–(A4), (A5′), *and* (A7). *Let $i \in \mathsf{Q}$.*

(i) *If $\lambda_i^* > 0$, then, as $t \to \infty$,*

$$t^{-\kappa_i} e^{-\lambda_i^* t} N_i(t) \xrightarrow{\text{a.s.}} \mathcal{N}_i := \frac{a_i}{\lambda_i^*}\mathcal{X}_i. \tag{11.2}$$

(ii) *If $\lambda_i^* = 0$, then, as $t \to \infty$,*

$$t^{-\kappa_i - 1} N_i(t) \xrightarrow{\text{a.s.}} \mathcal{N}_i := \frac{a_i}{\kappa_i + 1}\mathcal{X}_i. \tag{11.3}$$

Furthermore, if (A5) *or* (A8) *holds, then $\mathcal{N}_i > 0$ a.s.*

Theorem 11.2. *Let $(X_i(t))_{i\in\mathsf{Q}}$ be a discrete-time triangular Pólya urn and suppose that it satisfies* (A1)–(A5). *Alternatively, suppose that the urn satisfies* (A1)–(A4), (A5′), *and* (A6)–(A7), *and also either satisfies* (A8) *or is balanced. Let $i \in \mathsf{Q}$.*

(i) *If $\lambda_i^* > 0$, then as $n \to \infty$,*

$$\frac{N_{ni}}{n^{\lambda_i^*/\widehat{\lambda}} \log^{\gamma_i} n} \xrightarrow{\text{a.s.}} \widehat{\mathcal{N}}_i := \frac{a_i}{\lambda_i^*}\widehat{\mathcal{X}}_i. \tag{11.4}$$

(ii) *If $\lambda_i^* = 0$ and $\widehat{\lambda} > 0$, then as $n \to \infty$,*

$$\frac{N_{ni}}{\log^{\kappa_i + 1} n} \xrightarrow{\text{a.s.}} \widehat{\mathcal{N}}_i := \frac{a_i}{(\kappa_i + 1)\widehat{\lambda}}\widehat{\mathcal{X}}_i. \tag{11.5}$$

(iii) *If* $\lambda_i^* = \widehat{\lambda} = 0$*, then as* $n \to \infty$*,*

$$\frac{N_{ni}}{n^{(\kappa_i+1)/\widehat{\kappa}_0}} \xrightarrow{\text{a.s.}} \widehat{\mathcal{N}}_i := \frac{a_i}{\kappa_i + 1} \mathcal{X}_0^{-1/\widehat{\kappa}_0} \widehat{\mathcal{X}}_i. \tag{11.6}$$

Furthermore, if (A5) *or* (A8) *holds, then* $\widehat{\mathcal{N}}_i > 0$ *a.s.*

Remark 11.3. If we assume only (A1)–(A4), (A5′), and (A6)–(A7) in Theorem 11.2, then the results hold on the event $\{\mathcal{X}_0 > 0\}$, by the same proof and Theorem 8.5. △

We can also state the results as the following simple a.s. limit results for the ratios N_{ni}/X_{ni} and $N_i(t)/X_i(t)$; in particular, if $\lambda_i^* > 0$ (the main case), these ratios converge a.s. to a positive constant.

Theorem 11.4. *Let* $(X_i(t))_{i\in \mathsf{Q}}$ *be a discrete-time triangular Pólya urn and suppose that it satisfies* (A1)–(A5) *or (more generally)* (A1)–(A4), (A5′)*, and* (A7)–(A8)*. Let* $i \in \mathsf{Q}$*.*

(i) *If* $\lambda_i^* > 0$*, then*

$$\lim_{n\to\infty} \frac{N_{ni}}{X_{ni}} = \lim_{t\to\infty} \frac{N_i(t)}{X_i(t)} = \frac{a_i}{\lambda_i^*} \qquad \textit{a.s.} \tag{11.7}$$

(ii) *If* $\lambda_i^* = 0$*, then, as* $t \to \infty$*,*

$$\frac{N_i(t)}{tX_i(t)} \xrightarrow{\text{a.s.}} \frac{a_i}{\kappa_i + 1}. \tag{11.8}$$

(iii) *If* $\lambda_i^* = 0$ *and* $\widehat{\lambda} > 0$*, then, as* $n \to \infty$*,*

$$\frac{N_{ni}}{X_{ni} \log n} \xrightarrow{\text{a.s.}} \frac{a_i}{(\kappa_i + 1)\widehat{\lambda}}. \tag{11.9}$$

(iv) *If* $\lambda_i^* = 0$ *and* $\widehat{\lambda} = 0$*, then, as* $n \to \infty$*,*

$$\frac{N_{ni}}{X_{ni} n^{1/\kappa_0}} \xrightarrow{\text{a.s.}} \frac{a_i}{\kappa_i + 1} \mathcal{X}_0^{-1/\widehat{\kappa}_0}. \tag{11.10}$$

Note that the limit in all cases is a strictly positive constant, in case (iv) by Theorem 6.5 (and Remark 8.8).

Proof of Theorem 11.1. We use (as in the proof of the corresponding result in [27]) dummy balls similarly to the proof in Section 5, but now we use one dummy ball for each colour.

It suffices to consider one colour at a time, so we fix $i \in \mathsf{Q}$; we assume $a_i > 0$ since otherwise $N_{ni} = N_i(t) = 0$ a.s. and the results are trivial. Denote the urn by $\mathcal{U}$. We consider one new colour, which we denote by ι, and let $\mathsf{Q}^+ := \mathsf{Q} \cup \{\iota\}$ be the new set of colours. Balls of colour ι have activity $a_\iota := 0$ and are thus never drawn, and we let $\xi_{\iota j} := 0$ for all $j \in \mathsf{Q}^+$; we further let $\xi_{i\iota} := 1$ and $\xi_{j\iota} := 0$ for every $j \neq i$, and we start with $X_{0\iota} := 0$.

Consequently, the new urn, which we denote by $\mathcal{U}^+$, differs from the old one $\mathcal{U}$ only in that one additional ball of colour ι is added each time a ball of colour i is drawn. We may thus assume that the two urns are coupled such that they have the same $X_j(t)$ for all $j \in \mathsf{Q}$ and $t \geqslant 0$, and then

$$N_i(t) = X_\iota(t) \qquad \text{and} \qquad N_{ni} = X_{n\iota}. \tag{11.11}$$

The new urn $\mathcal{U}^+$ is also triangular and satisfies (A1)–(A5), or (A1)–(A4), (A5′), and (A7), if $\mathcal{U}$ does. In $\mathcal{U}^+$ we have $\lambda_\iota = 0$ and, using (2.8) and (2.9),

$$\lambda_\iota^* = \lambda_i^*, \tag{11.12}$$

$$\kappa_\iota = \begin{cases} \kappa_i, & \lambda_i^* > 0, \\ \kappa_i + 1, & \lambda_i^* = 0. \end{cases} \tag{11.13}$$

(i): By Theorem 4.1 or Theorem 8.6 applied to the new urn $\mathcal{U}^+$, we have (4.1) for the colour ι, and thus, using (11.11) and (11.12)–(11.13),

$$t^{-\kappa_i} e^{-\lambda_i^* t} N_i(t) = t^{-\kappa_\iota} e^{-\lambda_\iota^* t} X_\iota(t) \xrightarrow{\text{a.s.}} \mathcal{X}_\iota. \tag{11.14}$$

Furthermore, Lemma 3.5 or Lemma 8.11(a) applies to $\iota \in \mathsf{Q}^+$ (with i replaced by ι and j by i), and thus (3.35) yields, since $\lambda_\iota = 0 < \lambda_i^*$,

$$\mathcal{X}_\iota = \frac{a_i r_{i\iota}}{\lambda_i^* - \lambda_\iota} \mathcal{X}_i = \frac{a_i}{\lambda_i^*} \mathcal{X}_i. \tag{11.15}$$

Hence, (11.2) follows, with $\mathcal{N}_i := \mathcal{X}_\iota$.

(ii): Similar, now with $\kappa_\iota = \kappa_i + 1$ by (11.13), and using the second alternative in (3.35) which gives

$$\mathcal{N}_i := \mathcal{X}_\iota = \frac{a_i r_{i\iota}}{\kappa_\iota} \mathcal{X}_i = \frac{a_i}{\kappa_i + 1} \mathcal{X}_i. \tag{11.16}$$

The final sentence follows from Theorems 4.1 and 8.6, which yield $\mathcal{X}_\iota > 0$ a.s. □

Proof of Theorem 11.2. We add a dummy colour ι as in the proof of Theorem 11.1, and note that if the original urn $\mathcal{U}$ satisfies (A8) or is balanced, then the same holds for the new urn $\mathcal{U}^+$. (Note that dummy colours with activity 0 are ignored in Definition 10.1.)

By Theorem 1.8, Theorem 8.4, or Theorem 8.5 together with Lemma 10.9, the conclusions of Theorem 1.8 hold for $\mathcal{U}^+$, except that $\widehat{\mathcal{X}}_\iota = 0$ is possible unless we have (A5′) or (A8). (However, our assumptions yield $\mathcal{X}_0 > 0$ a.s. in all cases.)

We argue similarly to the proof of Theorem 11.1, now using (1.3) or (1.4) for the dummy colour ι in $\mathcal{U}^+$, together with (11.11) and (11.12)–(11.13). This yields a.s. convergence to $\widehat{\mathcal{N}}_i := \widehat{\mathcal{X}}_\iota$ in (11.4), (11.5), or (11.6) (depending on λ_i^* and $\widehat{\lambda}$); note that $\widehat{\lambda} \geqslant \lambda_i^*$, that (2.13) yields $\gamma_\iota = \gamma_i$ in (i) and $\gamma_\iota = \kappa_\iota = \kappa_i + 1$ in (ii), and that (2.12) shows that $\widehat{\kappa}_0$ is the same for $\mathcal{U}^+$ as for $\mathcal{U}$.

Finally, the formulas for $\widehat{\mathcal{N}}_i$ follow from (11.15)–(11.16) and (5.12) or (5.14). □

Alternatively, we may prove Theorem 11.2 from Theorem 11.1 by adding a dummy colur 0 as in Section 5. In any case, the proof is really based on adding two dummy colours 0 and ι to the continuous-time urn.

Proof of Theorem 11.4. The results for the continuous-time urn in (i) and (ii) follow by comparing the results of Theorem 11.1 with the limits for $X_i(t)$ in Theorem 4.1 or Theorem 8.6, recalling that $\mathcal{X}_i > 0$ a.s. as stated in Theorem 4.1. The result for N_{ni}/X_{ni} in (i) then follows by (11.1) and (2.19). Similarly, (11.9) and (11.10) follow from (11.8) and (5.9) or (5.13). (Alternatively, the results for N_{ni}/X_{ni} follow by comparing the results in Theorem 11.2 and Theorem 1.8.) □

Remark 11.5. In Theorem 11.4, if we assume only (A1)–(A4), (A5′), and (A7) but not (A8), then the conclusions hold on the event $\{\widehat{\mathcal{X}}_i > 0\}$ for $N_i(t)$ and on $\{\widehat{\mathcal{X}}_i > 0, \mathcal{X}_0 > 0\}$ for N_{ni}. △

12. Moment convergence

We have so far considered convergence a.s., which as always implies convergence in distribution. We consider in this section whether we also have convergence of moments, or equivalently convergence in L^p; recall the general fact that for a sequence of positive random variables converging a.s. (as we have here), convergence of the pth moment is equivalent to convergence in L^p, for any (real) $p > 0$, see e.g. [21, Theorem 5.5.2].

Unlike earlier sections, there seems to be an important difference between the discrete-time and continuous-time cases.

For a continuous-time urn, we will see below (Theorem 12.2) that we always have convergence in L^2 in Theorems 4.1 and 8.6; hence the mean and variance converge in these results. This extends to convergence in L^p for any $p > 2$, and thus convergence of any moment, assuming a corresponding moment condition for the replacements ξ_{ij} (Theorem 12.3).

The situation for discrete-time urns is more complicated. We will only consider balanced urns, and then prove L^2-convergence, and thus convergence of mean and variance; we also extend this to L^p and higher moments under the corresponding moment condition for the replacements ξ_{ij} (Theorem 12.5). It seems likely that this result extends to a wider class of triangular urns. However, it does *not* hold for all triangular urns. Example 14.2 gives a simple example where the a.s. limit does not have a finite mean, and thus we cannot have even L^1-convergence in Theorem 1.8.

Remark 12.1. The counterexample in Example 14.2 is rather special (a diagonal urn); [28, Theorem 1.6] shows that for a class of more typical unbalanced triangular urns, the a.s. limit in Theorem 1.8 has moments of all orders. This does not prove moment convergence, but we see no reason against it, and we conjecture that for these urns, and many others, we have convergence in L^p for all $p > 0$ (Problem 12.6). △

L^p-convergence or moment convergence (usually as part of a proof of convergence in distribution by the method of moments) have been proved earlier by different methods for some discrete-time Pólya urns, as far as we know all of them balanced. This includes balanced triangular urns with $q = 2$ or 3 and deterministic integer-valued replacements by Flajolet, Dumas and Puyhaubert [17], [43] (see Examples 14.3 and 14.11), and, for L^2 only, more general balanced triangular urns with deterministic replacements by Bose, Dasgupta, and Maulik [12] (Example 14.12). Further examples where moment convergence has been shown earlier are discussed in Examples 14.6, 14.7, 14.8, and 14.10.

12.1. **Continuous-time urns.**

Theorem 12.2. *In Theorems 4.1 and 8.6, the limit* (4.1) *holds also in* L^2*. In particular, mean and variance converge.*

Proof. Let $i \in \mathsf{Q}$. The proofs of Theorems 4.1 and 8.6 show that (3.83) holds. Thus, by the definitions (3.29)–(3.30),

$$\sup_{t\geqslant 1}\bigl|\widetilde{X}_i(t)\bigr|^2 \leqslant \bigl|2^{\kappa_i}\widetilde{X}_i^{**}\bigr|^2 \in L^1. \tag{12.1}$$

Hence, the collection $\{|\widetilde{X}_i(t)|^2 : t \geqslant 1\}$ is uniformly integrable, and consequently the a.s. convergence $\widetilde{X}_i(t) \to \mathcal{X}_i$ implies convergence in L^2 [21, Theorems 5.4.4 and 5.5.2]. □

Theorem 12.2 extends to L^p for all $p \geqslant 2$ as follows. The proof is based on the arguments in Section 3 combined with the Burkholder–Davis–Gundy inequalities which enable us to replace L^2 estimates by L^p estimates. The details are quite long, however, so we postpone them to Appendix B.

Theorem 12.3. *Let $p \geqslant 2$. In Theorems 4.1 and 8.6, assume also $\xi_{ij} \in L^p$ for all $i, j \in \mathsf{Q}$. Then, for every $i \in \mathsf{Q}$, the limit* (4.1) *holds also in L^p. Moreover, $\widetilde{X}_i^{**} \in L^p$.*

Remark 12.4. In Appendix B, we further show that Theorem 12.3 holds for any $p > 1$, also without the L^2 condition (A4). The proofs below then show that the same holds for all L^p results in this section. △

12.2. **Balanced discrete-time urns.**

Theorem 12.5. *Consider a balanced triangular urn statisfying* (A1)–(A5) *or* (A1)–(A4), (A5′), *and* (A7).

(i) *In Theorems 1.8 and 8.4–8.5, the limit* (1.3) *or* (1.4) *holds also in L^2. In particular, mean and variance converge.*
(ii) *Moreover, if $p \geqslant 2$ and $\xi_{ij} \in L^p$ for all $i, j \in \mathsf{Q}$, then, for every $i \in \mathsf{Q}$, the limit* (1.3) *or* (1.4) *holds also in L^p. Hence all moments of order $\leqslant p$ converge.*

As said above, this has been shown earlier in some special cases and examples, in particular [17], [43] ($q = 2, 3$), and [12] ($p = 2$); see also the examples in Section 14.

Proof. Part (i) is a special case of (ii), so we show only the latter.

Note first that (A6) holds by Lemma 10.2, and that Lemma 10.7 shows that $\mathcal{X}_0 > 0$ a.s.; thus the conclusion (1.3) or (1.4) of Theorem 1.8 holds by Theorem 1.8 or 8.5. (Or Theorem 8.4 when it is applicable.)

Let $c > 0$ be so small that $\mathbb{P}(X_0 > c) > \frac{1}{2}$. Our assumptions imply (as in the proof of Theorems 8.4 and 8.5) that (5.5) holds, by the proof in Section 5. This implies, by our choice of c,

$$\mathbb{P}\Big(\widehat{T}_n^{-\kappa_0} e^{-\widehat{\lambda}\widehat{T}_n} n > c\Big) > \frac{1}{2} \tag{12.2}$$

for all large n. By decreasing c (or just by ignoring some small n in the sequel), we may assume that (12.2) holds for all $n \geqslant 1$. Let $\mathcal{E}_n$ denote the event

$$\mathcal{E}_n := \big\{\widehat{T}_n^{-\kappa_0} e^{-\widehat{\lambda}\widehat{T}_n} n > c\big\}, \tag{12.3}$$

so that (12.2) reads $\mathbb{P}(\mathcal{E}_n) > \frac{1}{2}$.

Let $i \in \mathsf{Q}$. By Theorem 12.3, $\widetilde{X}_i^{**} \in L^p$. Hence, by (2.19) and (3.30),

$$Y_n := \left|\frac{X_{ni}}{\big(1 + \widehat{T}_n\big)^{\kappa_i} e^{\lambda_i^* \widehat{T}_n}}\right|^p = \left|\frac{X_i(\widehat{T}_n)}{\big(1 + \widehat{T}_n\big)^{\kappa_i} e^{\lambda_i^* \widehat{T}_n}}\right|^p \leqslant (\widetilde{X}_i^{**})^p \in L^1. \tag{12.4}$$

Consequently, the sequence Y_n is uniformly integrable. It follows from this and (12.2) that the sequence of conditioned random variables $(Y_n \mid \mathcal{E}_n)$ also is uniformly integrable. We consider two cases:

Case 1: $\widehat{\lambda} > 0$. In this case, Lemma 10.7 shows, using (5.3) and (2.13), that $\kappa_0 = \widehat{\kappa} = 0$ and $\gamma_i = \kappa_i$. Hence, the event $\mathcal{E}_n$ means $e^{-\widehat{\lambda}\widehat{T}_n} n > c$, and thus $\widehat{T}_n \leqslant t_n := \widehat{\lambda}^{-1}(\log n + C)$. Consequently, on the event $\mathcal{E}_n$ we have (for $n \geqslant 2$)

$$Z_n := \left|\frac{X_{ni}}{n^{\lambda_i^*/\widehat{\lambda}} \log^{\gamma_i} n}\right|^p = \left|\frac{X_{ni}}{n^{\lambda_i^*/\widehat{\lambda}} \log^{\kappa_i} n}\right|^p \leqslant C \left|\frac{X_{ni}}{t_n^{\kappa_i} e^{\lambda_i^* t_n}}\right|^p \leqslant C Y_n \tag{12.5}$$

and thus the uniform integrability of $(Y_n \mid \mathcal{E}_n)$ implies uniform integrability of $(Z_n \mid \mathcal{E}_n)$. However, by Lemma 10.4 and (12.3), X_{ni} is independent of the event $\mathcal{E}_n$; hence, so is Z_n and thus $(Z_n \mid \mathcal{E}_n) \overset{\mathrm{d}}{=} Z_n$. Consequently, the sequence Z_n is uniformly integrable.

In other words, the left-hand side of (1.3) is uniformly pth power integrable. Hence, the a.s. convergence in (1.3) implies convergence also in L^p.

Case 2: $\widehat{\lambda} = 0$. This is similar. In this case, (12.3) means $\widehat{T}_n \leqslant Cn^{1/\kappa_0}$. We now define

$$Z_n := \left| \frac{X_{ni}}{n^{\kappa_i/\widehat{\kappa}_0}} \right|^p, \tag{12.6}$$

and note again that Z_n is independent of $\mathcal{E}_n$. Since $0 \leqslant \lambda_i^* \leqslant \widehat{\lambda} = 0$ we have $\lambda_i^* = 0$, and $\widehat{\kappa}_0 = \kappa_0$ by (5.4); hence (12.6) and (12.4) show that on $\mathcal{E}_n$, we have $Z_n \leqslant CY_n$. Consequently, we have again $Z_n \overset{\mathrm{d}}{=} (Z_n \mid \mathcal{E}_n) \leqslant (CY_n \mid \mathcal{E}_n)$, and it follows that Z_n is uniformly integrable. Hence the a.s. convergence (1.4) holds also in L^p. ☐

As said above, Example 14.2 shows that Theorem 12.5 does not extend to all triangular urns. However, it seems likely that it extends to many unbalanced urns; we leave this as an open problem.

Problem 12.6. Find more general conditions (including also some unbalanced urns) for convergence in L^2 or L^p in Theorems 1.8 and 8.4.

12.3. **Moments for drawn colours.** The results above on convergence in L^2, and thus convergence of mean and variance, apply to the number of drawn balls with a given colour, N_{ni} and $N_i(t)$, since as shown in the proofs in Section 11, they can be regarded as $X_{n\iota}$ and $X_\iota(t)$ for an extended urn with a dummy colour ι added. Hence we obtain:

Theorem 12.7. *In Theorem 11.1, the a.s. limit* (11.2) *or* (11.3) *holds also in* L^2. *Moreover, if* $p \geqslant 2$ *and* $\xi_{ij} \in L^p$ $\forall i, j \in \mathsf{Q}$*, then the limit holds also in* L^p.

Proof. By Theorem 12.2 or 12.3 applied to $X_\iota(t)$. ☐

Theorem 12.8. *In Theorem 11.2, if the urn is balanced, then the a.s. limit* (11.4)*,* (11.5)*, or* (11.6) *holds also in* L^2. *Moreover, if* $p \geqslant 2$ *and* $\xi_{ij} \in L^p$ $\forall i, j \in \mathsf{Q}$*, then the limit holds also in* L^p.

Proof. By Theorem 12.5 applied to $X_{n\iota}$. ☐

13. Rates of convergence?

For classical Pólya urns (Example 14.1), the rate of convergence for convergence in distribution in (14.4) or (14.5) has been studied, for several different metrics; see [29] and the references there. As noted in [29, Remark 1.4], the rate of a.s. convergence is slower, and is the same as in the law of large numbers for i.i.d. Bernoulli variables, which is given by the law of iterated logarithm.

For other triangular urns, we are not aware of any similar results on rates of convergence; however, [17] gives upper bounds for the rate of convergence of moments and in a local limit theorem, for some balanced triangular urns with deterministic replacements. (Irreducible, and thus non-triangular, balanced urns with $q = 2$ and deterministic replacements are studied in [34].)

Problem 13.1. Study rates of convergence in e.g. (1.3) and (4.1), both for the a.s. convergence and for convergence in distribution.

Note that this problem is closely related to the problem studying fluctuations from the limit, mentioned in Remark 1.9.

14. Examples

We consider several examples, many of which have been treated earlier from different perspectives. The purpose is to illustrate both the theorems above and (some of) their relations to earlier literature. We generally label the colours by $1, \dots, q$, and then assume $\xi_{ij} = 0$ when $i < j$. We assume that the initial composition $\mathbf{X}_0$ is deterministic. We write for convenience $x_i := X_{i0} = X_i(0)$ and $\mathbf{x} := (x_i)_1^q = \mathbf{X}_0$. We denote the total number of balls in the urn after n draws by $|\mathbf{X}_n| := \sum_{i=1}^q X_{ni}$.

When $q = 2$ we sometimes also call the colours *white* and *black* (in this order; thus a black draw may add only black balls: "there is no escape from a black hole"). We then may write $W_n := X_{n1}$, $B_n := X_{n2}$, $W(t) := X_1(t)$, $B(t) := X_2(t)$, $w_0 := x_1 = X_{10}$, $b_0 := x_2 = X_{20}$.

We usually describe the urns using the replacement matrix $(\xi_{ij})_{i,j=1}^q$ where the rows are the replacement vectors. (See Remark 1.4.)

In all our examples, all activities $a_i = 1$. Thus

$$\lambda_i = r_{ii} = \mathbb{E}\,\xi_{ii}, \qquad i \in \mathsf{Q} = \{1, \dots, q\}. \tag{14.1}$$

Example 14.1. The classical Pólya urn has balls of q colours; when a ball is drawn it is replaced together with a fixed number $b > 0$ balls of the same colour. Hence, the replacement matrix is deterministic and diagonal, with entries b on the diagonal. This urn is obviously balanced, and we have $\lambda_i = \lambda_i^* = b$ and $\kappa_i = 0$ for every colour i.

This urn model was studied (for $q = 2$) already by Markov [39], Eggenberger and Pólya [16] and Pólya [41]. See also e.g. Johnson and Kotz [31, Chapter 4] and Mahmoud [38].

For this urn (as for any diagonal urn), in the continuous-time version, the different colours evolve independently, and each colour is version of the Yule process. More precisely, $X_i(t)/b$ is a Yule process started with x_i/b individuals, where each individual gets children at rate b; thus $X_i(t/b)/b$ is a Yule process with the standard rate 1. (This is a classical branching process if x_i/b is an integer, and in general a CB process.) It is well-known that in this case $e^{-bt}X_i(t)/b \xrightarrow{\text{d}} \Gamma(x_i/b, 1)$ and thus

$$e^{-bt}X_i(t) \xrightarrow{\text{a.s.}} \mathcal{X}_i \in \Gamma(x_i/b, b); \tag{14.2}$$

furthermore, $\mathcal{X}_1, \dots, \mathcal{X}_q$ are independent, since the processes $X_i(t)$ are independent. This is an example of Theorem 4.1. Moreover, (5.11)–(5.12), or (5.16) where now $\mathsf{Q}_* = \mathsf{Q}$, show together with (5.17) that

$$\frac{X_{in}}{n} \xrightarrow{\text{a.s.}} \widehat{\mathcal{X}}_i := b\frac{\mathcal{X}_i}{\sum_{j=1}^q \mathcal{X}_j}. \tag{14.3}$$

It follows that the vector of proportions converges:

$$\frac{\mathbf{X}_n}{|\mathbf{X}_n|} \xrightarrow{\text{a.s.}} \frac{1}{b}\widehat{\boldsymbol{\mathcal{X}}} := \frac{1}{b}\big(\widehat{\mathcal{X}}_1, \dots, \widehat{\mathcal{X}}_q\big) = \frac{(\mathcal{X}_1, \dots, \mathcal{X}_q)}{\sum_{j=1}^q \mathcal{X}_j}, \tag{14.4}$$

where, as a consequence of (14.2), the limit vector $b^{-1}\widehat{\boldsymbol{\mathcal{X}}}$ has a Dirichlet distribution with parameter $\mathbf{x}/b$. In particular, each marginal converges a.s. to a Beta distributed variable:

$$\frac{X_{ni}}{|\mathbf{X}_n|} \xrightarrow{\text{a.s.}} b^{-1}\widehat{\mathcal{X}}_i \sim B\Bigl(\frac{x_i}{b}, \sum_{j\neq i}\frac{x_j}{b}\Bigr). \tag{14.5}$$

These results are all well known; the limit (14.5) with convergence in distribution was shown for $q = 2$ already in [39] and [41], and for general q in [10] (in a special case) and [2]; see also [31, Section 6.3.3] and [38]. Furthermore, a.s. convergence has been shown by a number of methods, for example in [10] and [2]. △

Example 14.2. A *diagonal* Pólya urn has $\xi_{ij} = 0$ for $i \neq j$; in other words, all added balls have the same colour as the drawn ball. This is a generalization of the classical Pólya urn in Example 14.1, but now the diagonal elements ξ_{ii} can be random, and they may have different distributions. A.s. convergence for this urn has been shown, under weak technical conditions, by Athreya [2]; see also Aguech [1, Theorem 4].

Consider for simplicity the case when the replacements are deterministic, and assume to avoid trivialities that $r_{ii} = \xi_{ii} > 0$ for every $i \in \mathsf{Q}$.

As in Example 14.1, in the continuous-time urn, the colours evolve as independent Yule processes, now with possibly different rates $\lambda_i = \xi_{ii}$. Hence, generalizing (14.2),

$$e^{-\lambda_i t}X_i(t) \xrightarrow{\text{a.s.}} \mathcal{X}_i \in \Gamma(x_i/\lambda_i, \lambda_i), \tag{14.6}$$

with all $\mathcal{X}_i$ independent.

Consider the simplest case: $q = 2$, and assume $\lambda_1 = \alpha$ and $\lambda_2 = \delta$ with $\alpha > \delta > 0$. Then $\lambda_i^* = \lambda_i$, $\widehat{\lambda} = \alpha$, $\kappa_i = \widehat{\kappa} = \gamma_i = 0$ $(i = 1, 2)$. It follows from (5.11)–(5.12) and (5.17) (or directly from (14.6)) that

$$n^{-\delta/\alpha}X_{n2} \xrightarrow{\text{a.s.}} \widehat{\mathcal{X}}_2 = \alpha^{\delta/\alpha}\frac{\mathcal{X}_2}{\mathcal{X}_1^{\delta/\alpha}}. \tag{14.7}$$

Note that if $\mathcal{X} \sim \Gamma(a, b)$, then its moments (for arbitrary real r) are given by

$$\mathbb{E}\,\mathcal{X}^r = \begin{cases} b^r\Gamma(a+r)/\Gamma(a) < \infty, & -a < r < \infty,\\ \infty, & r \leqslant -a. \end{cases} \tag{14.8}$$

In particular, since $\mathcal{X}_1$ and $\mathcal{X}_2$ are independent, it follows from (14.6)–(14.8) that for $r > 0$,

$$\mathbb{E}\,\widehat{\mathcal{X}}_2^r < \infty \iff \mathbb{E}\,\mathcal{X}_1^{-r\delta/\alpha} < \infty \iff r\delta/\alpha < x_1/\alpha \iff r < x_1/\delta. \tag{14.9}$$

Consequently, $\widehat{\mathcal{X}}_2$ does not have finite moments of all orders. In particular, we cannot always have (finite) moment convergence in (14.7). Taking, for example, $\alpha = 2$, $\delta = 1$, and $x_1 = x_2 = 1$ we see that not even the mean $\mathbb{E}\,\widehat{\mathcal{X}}_2$ is finite; hence we cannot have convergence in L^1 or L^2 in Theorem 1.8.

As far as we know, it is an open problem to find asymptotics of moments $\mathbb{E}\,X_{n2}^r$ for general $r > 0$ in this (simple) example. △

Example 14.3. Consider a two-colour urn with a deterministic replacement matrix

$$\begin{pmatrix} \delta & \gamma\\ 0 & \alpha \end{pmatrix}. \tag{14.10}$$

(We have chosen a notation agreeing with [28], although colours are taken in different order there and thus the matrices are written differently.) This urn (and special cases of it) have been studied in many papers; in particular, [28] gives a detailed study of limits in distribution. The balanced case $\alpha = \delta + \gamma$ with integers α, γ, δ is studied by very different methods (generating functions) in [43] and [17]. A.s. convergence has been shown in special cases in [18; 19; 11; 12] ($\alpha = \delta + \gamma$), and [1] ($\alpha \geqslant \delta$).

Suppose that $\delta > 0$, $\gamma > 0$, $w_0 = x_1 > 0$, and $b_0 = x_2 \geqslant 0$; suppose also either $\alpha \geqslant 0$, or $\alpha = -1$ together with $\gamma \in \mathbb{Z}_+$ and $x_2 \in \mathbb{Z}_{\geqslant 0}$. Then the urn satisfies (A1)–(A5) if $\alpha \geqslant 0$, and (A1)–(A4), (A5′), and (A7)–(A8) for all α. Hence, Theorems 1.8 and 4.1 apply if $\alpha \geqslant 0$, and Theorems 8.4 and 8.6 apply for any α; consequently, the conclusions of Theorems 1.8 and 4.1 hold for all cases.

We have $\lambda_1 = \delta$ and $\lambda_2 = \alpha$. Furthermore, since $\gamma > 0$, we have $1 \to 2$. (Thus 1 is the only minimal colour.) Hence, $\lambda_1^* := \delta$, $\lambda_2^* := \alpha \vee \delta$ and thus $\widehat{\lambda} = \lambda_2^* = \alpha \vee \delta$; furthermore $\kappa_1 = 0$ while $\kappa_2 = 1$ when $\alpha = \delta$ and $\kappa_2 = 0$ otherwise. We consider several cases.

Case 1, $\alpha < \delta$:

Then $\lambda_1^* = \lambda_2^* = \widehat{\lambda} = \delta > 0$; furthermore, $\kappa_1 = \kappa_2 = 0 = \widehat{\kappa}$, and (2.13) yields $\gamma_1 = \gamma_2 = 0$. Consequently, Theorem 1.8(i) or Theorem 8.4 yields

$$\frac{X_{ni}}{n} \xrightarrow{\text{a.s.}} \widehat{\mathcal{X}}_i, \qquad i = 1, 2. \tag{14.11}$$

Furthermore, 1 is the only leader, and thus Theorem 7.2 shows that $\widehat{\mathcal{X}}_1$ and $\widehat{\mathcal{X}}_2$ are constants. To find them, we can use Lemma 6.7. By Lemma 6.1 (simplifying the notation), $\mathcal{X}_i = c_i \mathcal{X}_1$, where obviously $c_1 = 1$, and Lemma 6.7 gives the eigenvalue equation

$$(c_1, c_2) \begin{pmatrix} \delta & \gamma \\ 0 & \alpha \end{pmatrix} = \delta (c_1, c_2), \tag{14.12}$$

i.e., $\gamma + \alpha c_2 = \delta c_2$, with the solution $c_2 = \gamma/(\delta - \alpha)$. In other words,

$$\mathcal{X}_2 = \frac{\gamma}{\delta - \alpha} \mathcal{X}_1. \tag{14.13}$$

This follows also directly from Lemma 3.5 and (3.35).

If we add a dummy colour 0 as in Section 5, then (5.17) and (14.13) yield

$$\mathcal{X}_0 = \delta^{-1}(\mathcal{X}_1 + \mathcal{X}_2) = \frac{1}{\delta} \frac{\gamma + \delta - \alpha}{\delta - \alpha} \mathcal{X}_1. \tag{14.14}$$

Hence, by (5.12),

$$\widehat{\mathcal{X}}_1 = \frac{\mathcal{X}_1}{\mathcal{X}_0} = \frac{\delta(\delta - \alpha)}{\gamma + \delta - \alpha}, \tag{14.15}$$

$$\widehat{\mathcal{X}}_2 = \frac{\mathcal{X}_2}{\mathcal{X}_0} = \frac{\delta\gamma}{\gamma + \delta - \alpha}. \tag{14.16}$$

Consequently, as $n \to \infty$ we have

$$\frac{X_{n1}}{n} \xrightarrow{\text{a.s.}} \widehat{\mathcal{X}}_1 = \frac{\delta(\delta - \alpha)}{\gamma + \delta - \alpha}, \tag{14.17}$$

$$\frac{X_{n2}}{n} \xrightarrow{\text{a.s.}} \widehat{\mathcal{X}}_2 = \frac{\delta\gamma}{\gamma + \delta - \alpha}. \tag{14.18}$$

This is in agreement with [28, Theorem 1.3(i)-(iii) and Lemma 1.2], which give the asymptotic distribution of the difference $X_{ni} - n\widehat{\mathcal{X}}_i$ divided by the correct normalization factor, which implies (and is much more precise than) convergence in probability in (14.17)–(14.18). (The normalization factor is $n^{1/2}$ for $\alpha < \delta/2$, $(n \log n)^{1/2}$ for $\alpha = \delta/2$ and $n^{\alpha/\delta}$ for $\delta/2 < \alpha < \delta$. Moreover, the distribution is asymptotically normal for $\alpha \leqslant \delta/2$, but not for $\delta/2 < \alpha < \delta$. See [28] for details.)

Case 2, $\alpha = \delta$:

Then $\lambda_1^* = \lambda_2^* = \widehat{\lambda} = \delta > 0$; furthermore, $\kappa_1 = 0$ and $\kappa_2 = 1 = \widehat{\kappa}$, and thus (2.13) yields $\gamma_1 = -1$ and $\gamma_2 = 0$. Consequently, Theorem 1.8(i) yields

$$\frac{X_{n1}}{n/\log n} \xrightarrow{\text{a.s.}} \widehat{\mathcal{X}}_1, \tag{14.19}$$

$$\frac{X_{n2}}{n} \xrightarrow{\text{a.s.}} \widehat{\mathcal{X}}_2. \tag{14.20}$$

Furthermore, also in this case, 1 is the only leader, and thus Theorem 7.2 shows that $\widehat{\mathcal{X}}_1$ and $\widehat{\mathcal{X}}_2$ are constants. However, unlike Case 1, $\lambda_2 = \lambda_2^*$ and thus 2 is now a subleader. Again, Lemma 6.1 shows that $\mathcal{X}_2 = c_2\mathcal{X}_1$, where (6.20) in Lemma 6.7 immediately yields

$$c_2 = \frac{a_1 r_{12}}{\kappa_2} c_1 = \gamma. \tag{14.21}$$

Furthermore, (5.15) now yields $\mathsf{Q}_* = \{2\}$, and thus (5.17) yields

$$\mathcal{X}_0 = \widehat{\lambda}^{-1}\mathcal{X}_2 = \delta^{-1}\mathcal{X}_2. \tag{14.22}$$

Consequently, recalling (5.12), the limits in (14.19)–(14.20) are

$$\widehat{\mathcal{X}}_1 = \widehat{\lambda}\frac{\mathcal{X}_1}{\mathcal{X}_0} = \delta^2\frac{\mathcal{X}_1}{\mathcal{X}_2} = \frac{\delta^2}{c_2} = \frac{\delta^2}{\gamma}, \tag{14.23}$$

$$\widehat{\mathcal{X}}_2 = \frac{\mathcal{X}_2}{\mathcal{X}_0} = \delta. \tag{14.24}$$

This is in agreement with the result on the asymptotic distribution in [28, Theorem 1.3(iv) and Lemma 1.2], which implies convergence in probability in (14.19)–(14.20).

Case 3, $\alpha > \delta$:

Then $0 < \lambda_1^* = \delta < \lambda_2^* = \alpha = \widehat{\lambda}$; furthermore, $\kappa_1 = \kappa_2 = 0 = \widehat{\kappa}$, and thus $\gamma_1 = \gamma_2 = 0$. Consequently, Theorem 1.8(i) yields (see also [1])

$$\frac{X_{n1}}{n^{\delta/\alpha}} \xrightarrow{\text{a.s.}} \widehat{\mathcal{X}}_1, \tag{14.25}$$

$$\frac{X_{n2}}{n} \xrightarrow{\text{a.s.}} \widehat{\mathcal{X}}_2. \tag{14.26}$$

In this case, both colours 1 and 2 are leaders; hence Theorem 6.4 shows that $\mathcal{X}_1$ and $\mathcal{X}_2$ are absolutely continuous, also jointly. Furthermore, Theorem 7.1(ii) shows that $\widehat{\mathcal{X}}_1$ is absolutely continuous, while Theorem 7.2 shows that $\widehat{\mathcal{X}}_2$ is deterministic. We have again $\mathsf{Q}_* = \{2\}$, which by (5.17) now yields

$$\mathcal{X}_0 = \widehat{\lambda}^{-1}\mathcal{X}_2 = \alpha^{-1}\mathcal{X}_2. \tag{14.27}$$

Hence, (5.12) yields

$$\widehat{\mathcal{X}}_1 = \frac{\mathcal{X}_1}{\mathcal{X}_0^{\delta/\alpha}} = \alpha^{\delta/\alpha}\frac{\mathcal{X}_1}{\mathcal{X}_2^{\delta/\alpha}}, \tag{14.28}$$

$$\widehat{\mathcal{X}}_2 = \frac{\mathcal{X}_2}{\mathcal{X}_0} = \alpha. \tag{14.29}$$

However, the formula (14.28) for $\widehat{\mathcal{X}}_1$ does not seem to be of much use to find the distribution of $\widehat{\mathcal{X}}_1$. This distribution was found by other methods in [28, Theorem 1.3(v)], which yields convergence in distribution of $X_{n1}/n^{\delta/\alpha}$; the a.s. convergence in (14.25) is a stronger result, and the distribution of $\widehat{\mathcal{X}}_1$ is thus the limit distribution found in [28]. This limit distribution is characterized in [28], but no simple form is known in general; it is shown in [28, Theorem 1.6] that $\widehat{\mathcal{X}}_1$ has moments of all orders, and a complicated integral formula is given for these moments. Except in the balanced case below, it is, as far as we know, an open problem whether moments converge in (14.25) (to these limits) or not.

Case 4, $\alpha = \delta + \gamma$:

The balanced case of the two-colour urn studied here is $\alpha = \delta + \gamma$; by our assumption $\gamma > 0$, this is a special case of Case 3, and thus (14.25)–(14.29) hold. In this special case, $\widehat{\mathcal{X}}_1$ can be characterized by its moments, for which there is a simple formula, see (for integers $\alpha, \gamma, \delta, r$) [43, Theorem 2.9], [17, Proposition 17], and (for the general case) [28, Theorem 1.7]:

$$\mathbb{E}\,\widehat{\mathcal{X}}_1^r = \delta^r \frac{\Gamma\big((x_1+x_2)/\alpha\big)\Gamma(x_1/\delta + r)}{\Gamma(x_1/\delta)\Gamma\big((x_1+x_2+r\delta)/\alpha\big)}, \qquad r > 0. \tag{14.30}$$

It follows that the moment generating function $\mathbb{E}\,e^{t\widehat{\mathcal{X}}_1}$ is finite for all real t, and thus the moments (14.30) (even for integer r) determine the distribution. Moreover, Theorem 12.5 shows that all moments converge in (14.25) (to the limits (14.30)); this was earlier shown in [17; 43] in the case that the replacements δ, γ, α are integers.

If we further assume $x_2 = 0$, so we start only with white balls, then $\widehat{\mathcal{X}}_1$ has a density function that can be expressed using the density function of a Mittag-Leffler distribution with parameter δ/α, or a δ/α-stable distribution, see [28, Theorem 1.8] and [17, Proposition 16].

The continuous-time processes $W(t)$ and $B(t)$ are studied by [13], which includes our Theorem 4.1 for this balanced urn. △

Example 14.4. Generalizing Example 14.3, consider a general triangular two-colour urn $\mathcal{U}$ with random replacement matrix

$$\begin{pmatrix} \xi_{11} & \xi_{12} \\ 0 & \xi_{22} \end{pmatrix}. \tag{14.31}$$

Such urns have been studied by Aguech [1], who proved (among other results) the existence of a.s. limits under some assumptions (including our (A5), $\mathbb{E}\,\xi_{22} \geqslant \mathbb{E}\,\xi_{11}$, and an unnecessary independence assumption). We extend this result as follows.

Assume $\xi_{11}, \xi_{12}, \xi_{22} \in L^2$ (Condition (A4)), and let $\delta := \mathbb{E}\,\xi_{11}$, $\gamma := \mathbb{E}\,\xi_{12}$, $\alpha := \mathbb{E}\,\xi_{22}$, so that $\left(\begin{smallmatrix} \delta & \gamma \\ 0 & \alpha \end{smallmatrix}\right)$ is the mean replacement matrix. Suppose that (A5′) holds, that $\xi_{11} \geqslant 0$ a.s., and that, as in Example 14.3, $\delta > 0$, $\gamma > 0$, $w_0 = x_1 > 0$, and $b_0 = x_2 \geqslant 0$. Then the urn satisfies (A1)–(A5) if $\xi_{22} \geqslant 0$ a.s., and (A1)–(A4), (A5′), and (A7)–(A8) in any case. Hence, Theorems 8.4 and 8.6 apply, and thus the conclusions of Theorems 1.8 and 4.1 hold for all α.

As in Section 9, we let $\mathcal{U}'$ denote the mean urn with replacement matrix $\left(\begin{smallmatrix} \delta & \gamma \\ 0 & \alpha \end{smallmatrix}\right)$; this urn is of the type in Example 14.3. As discussed in Section 9, all parameters $\lambda_i, \lambda_i^*, \widehat{\lambda}, \dots$ are the same for $\mathcal{U}$ and $\mathcal{U}'$, and thus all results are qualitatively the same for the two urns. Hence, all results in Example 14.3 hold for the urn $\mathcal{U}$ with random replacements too, except any assertions on the precise distributions of the limits (including the moment formula (14.30), which as shown in Example 14.5 below is *not* valid in general for random replacements). Note, however, that by Theorem 9.1, in all cases in Example 14.3 with a deterministic limit $\widehat{\mathcal{X}}_i$ for the mean urn, we have the same limit for the urn $\mathcal{U}$. △

Example 14.5. Consider a two-colour urn $\mathcal{U}$ with the random replacement matrix

$$\begin{pmatrix} \xi & 1-\xi \\ 0 & 1 \end{pmatrix}, \tag{14.32}$$

where $\xi \in \mathrm{Be}(p)$ for some $p \in (0,1)$. In other words, if we draw a white ball, we add another ball that is white with probability p, and otherwise black; if we draw a black ball we always add another black ball. (As usual, we also always return the drawn ball.) Note that this urn is balanced, in spite of replacements being random.

This urn appears in several applications, see Examples 14.6 and 14.10 for two of them.

This urn is a special case of Example 14.4, and a.s. convergence for the urn follows from Theorem 1.8. Moment convergence follows from Theorem 12.5.

The mean replacement matrix is

$$(r_{ij})_{i,j=1}^2 = \begin{pmatrix} p & 1-p \\ 0 & 1 \end{pmatrix}. \tag{14.33}$$

By Section 9, the asymptotic behaviour of the urn $\mathcal{U}$ is qualitatively the same as for the mean urn $\mathcal{U}'$ with the replacement matrix (14.33), which is an instance of Example 14.3, more precisely the balanced Case 4. In particular, since (14.29) shows that $\widehat{\mathcal{X}}_2 = \alpha = 1$ is constant for the mean urn $\mathcal{U}'$, the same holds for the original urn $\mathcal{U}$ by Theorem 9.1.

We may also relate the urn $\mathcal{U}$ to the mean urn $\mathcal{U}'$ in another, more direct, way. Conditioned on the contents (X_{n1}, X_{n2}) of the urn at time n, the probability that the next added ball is white is, letting ζ be the colour of the drawn ball,

$$\mathbb{P}\big(\zeta = 1 \mid X_{n1}, X_{n2}\big) \cdot p + \mathbb{P}\big(\zeta = 2 \mid X_{n1}, X_{n2}\big) \cdot 0 = \frac{pX_{n1}}{X_{n1} + X_{n2}}. \tag{14.34}$$

Hence, if we define

$$Y_{n1} := pX_{n1}, \tag{14.35}$$

$$Y_{n2} := (1-p)X_{n1} + X_{n2}, \tag{14.36}$$

and note that $Y_{n1} + Y_{n2} = X_{n1} + X_{n2}$, we see from (14.34) that we may regard the *added* ball in $\mathcal{U}$ as the *drawn* ball in an urn with composition $\mathbf{Y}_n = (Y_{n1}, Y_{n2})$. Adding a white ball to (X_{n1}, X_{n2}) (i.e., increasing X_{n1} by 1) means by (14.35)–(14.36) adding $(p, 1-p)$ to (Y_{n1}, Y_{n2}), while adding a black ball to (X_{n1}, X_{n2}) means adding $(0,1)$ to (Y_{n1}, Y_{n2}). Consequently, the stochastic process $(\mathbf{Y}_n)_{n\geqslant 0}$ describes a Pólya urn with the replacement matrix $\left(\begin{smallmatrix} p & 1-p \\ 0 & 1 \end{smallmatrix}\right)$, which is the same as (14.33) for the mean urn $\mathcal{U}'$ above. Note, however, that the initial conditions now are, by (14.35)–(14.36),

$$y_1 = px_1, \qquad y_2 = (1-p)x_1 + x_2. \tag{14.37}$$

By Example 14.3, we have

$$Y_{n1}/n^p \xrightarrow{\text{a.s.}} \widehat{\mathcal{Y}}_1 \tag{14.38}$$

where the limit by (14.30) has moments, recalling (14.37),

$$\mathbb{E}\,\widehat{\mathcal{Y}}_1^r = p^r \frac{\Gamma(y_1+y_2)\,\Gamma(y_1/p+r)}{\Gamma(y_1/p)\,\Gamma(y_1+y_2+rp)} = p^r \frac{\Gamma(x_1+x_2)\,\Gamma(x_1+r)}{\Gamma(x_1)\,\Gamma(x_1+x_2+rp)}. \tag{14.39}$$

Hence, by (14.35), we have in the urn $\mathcal{U}$ with replacements (14.32)

$$X_{n1}/n^p \xrightarrow{\text{a.s.}} \widehat{\mathcal{X}}_1 \tag{14.40}$$

with $\widehat{\mathcal{X}}_1 = p^{-1}\widehat{\mathcal{Y}}_1$, and thus

$$\mathbb{E}\,\widehat{\mathcal{X}}_1^r = p^{-r}\,\mathbb{E}\,\widehat{\mathcal{Y}}_1^r = \frac{\Gamma(x_1+x_2)\,\Gamma(x_1+r)}{\Gamma(x_1)\,\Gamma(x_1+x_2+rp)}, \qquad r \geqslant 0. \tag{14.41}$$

By comparing (14.40) with (14.30) for the mean urn, we see that the means $\mathbb{E}\,\widehat{\mathcal{X}}_1$ are the same for the two urns, while for the second moment, (14.41) and (14.30) yield, for the urn $\mathcal{U}$ and its mean urn $\mathcal{U}'$, respectively,

$$\mathbb{E}\,\widehat{\mathcal{X}}_1^2 = x_1(x_1+1)\frac{\Gamma(x_1+x_2)}{\Gamma(x_1+x_2+2p)}, \quad \mathbb{E}\,\widehat{\mathcal{X}}_1^2 = x_1(x_1+p)\frac{\Gamma(x_1+x_2)}{\Gamma(x_1+x_2+2p)}. \tag{14.42}$$

The variance is thus larger for the urn $\mathcal{U}$ with random replacement. (Perhaps not surprisingly.) This shows that an urn and its mean urn in general have different asymptotic distributions, although the qualitative behaviour is the same as shown in Section 9. △

Example 14.6. One example where the urn in Example 14.5 with replacement matrix (14.32) appears is that it describes the size of the root cluster for bond percolation (with parameter p) on the random recursive tree (with vertices in the root cluster coloured white and all other vertices black, and the initial vector $(1,0)$). (See the argument in the generalization Example 14.7 below. The root cluster is studied by other methods in [6], [35], [5], [14].) In this case we obtain (14.40) where (14.41) yields

$$\mathbb{E}\,\widehat{\mathcal{X}}_1^r = \frac{\Gamma(r+1)}{\Gamma(1+rp)}, \qquad r \geqslant 0, \tag{14.43}$$

which means that $\widehat{\mathcal{X}}_1$ has a Mittag-Leffler distribution; this was proved by Baur and Bertoin [6] (using other methods). △

Example 14.7. Baur [5] and Desmarais, Holmgren, and Wagner [14] considered (among other things) the root cluster in bond percolation on a preferential attachment tree, generalizing Example 14.6. The tree is defined as follows, for a real parameter α. Construct the rooted tree $\mathcal{T}_n$ with n vertices recursively, starting with $\mathcal{T}_1$ being just the root and adding vertices one by one; each new vertex is attached to a parent v chosen among existing vertices with probability proportional to $\alpha d(v)+1$, where $d(v)$ is the current outdegree of v. We also perform bond percolation, and let each edge by *active* with probability $p \in (0,1)$, independently of all other edges. (Note that both active and passive edges are counted in the outdegree.) A vertex is active if it is connected to the root by a path of active edges. Let Z_n be the number of active vertices in $\mathcal{T}_n$.

The case $\alpha = 0$ gives the random recursive tree in Example 14.6. We consider here the case $\alpha \geqslant 0$ (as assumed in [5]), and study the modifications for $\alpha < 0$ in the following example.

We model this process by an urn with two colours, where each vertex v contributes $\alpha d(v)+1$ balls, which are white if the urn is active and black otherwise. To find the parent of the next vertex corresponds to drawing a ball from the urn; if the ball is white then the parent is active and the new vertex becomes active with probability p (and otherwise passive); if the ball is black then the parent is passive and the new vertex always becomes passive. Since the outdegree of the parent increases by 1, we add α balls of the same colour as the drawn ball, plus one ball for the new vertex which is white if the new vertex is active, i.e., with probability p if the drawn ball is white, and otherwise black. Hence, the urn is a Pólya urn with random replacement matrix

$$\begin{pmatrix} \alpha+\xi & 1-\xi \\ 0 & \alpha+1 \end{pmatrix}, \tag{14.44}$$

where $\xi \in \mathrm{Be}(p)$. We start with 1 active vertex, and thus the urn starts with a single white ball, i.e., $\mathbf{x}=(1,0)$.

If $\alpha=0$ (the random recursive tree), we get again (14.32), as discussed in Example 14.6. In general, unlike Example 14.6, the number of active vertices Z_n is not directly reflected by the contents of the urn. However, if N_{n1} is the number of drawn white balls, then this is the number of vertices that get an active parent; hence N_{n1} is the total outdegree of the active vertices, and therefore the number of white balls in the urn is

$$W_n = X_{n1} = \alpha N_{n1} + Z_n. \tag{14.45}$$

The replacement matrix (14.44) shows that the urn is of the type in Example 14.4; furthermore, it is balanced with balance $\alpha+1$. We have $\lambda_1^* = \lambda_1 = \mathbb{E}\,\xi_{11} = \alpha+p$ and $\widehat{\lambda} = \lambda_2 = \alpha+1$; further $\kappa_1=\kappa_2=0$. Theorem 1.8 yields

$$\frac{X_{n1}}{n^{(\alpha+p)/(\alpha+1)}} \xrightarrow{\text{a.s.}} \widehat{\mathcal{X}}_1. \tag{14.46}$$

Moreover, Theorem 11.4 yields

$$\frac{N_{n1}}{X_{n1}} \xrightarrow{\text{a.s.}} \frac{1}{\lambda_1^*} = \frac{1}{\alpha+p}, \tag{14.47}$$

and thus (14.45) yields

$$\frac{Z_n}{X_{n1}} = \frac{X_{n1}-\alpha N_{n1}}{X_{n1}} \xrightarrow{\text{a.s.}} 1-\frac{\alpha}{\alpha+p} = \frac{p}{\alpha+p}. \tag{14.48}$$

Hence, (14.46) yields

$$\frac{Z_n}{n^{(\alpha+p)/(\alpha+1)}} \xrightarrow{\text{a.s.}} \mathcal{Z} := \frac{p}{\alpha+p}\widehat{\mathcal{X}}_1, \tag{14.49}$$

where the limit is in $(0,\infty)$ a.s. The limits (14.46) and (14.49) hold also in L^r for any $r<\infty$ by Theorems 12.5 and 12.8; hence all moments converge.

This complements Baur [5, Proposition 4.1], who shows L^2-convergence in (14.49) and gives the first two moments of the limit (by different but related methods), and Desmarais, Holmgren, and Wagner [14] who prove convergence of all moments in (14.49) and give a recursion for the moments of the limit. (The distribution of the limit is not known explicitly.) △

Example 14.8. In Example 14.7, we assumed $\alpha \geqslant 0$. However, the results extend easily to the case $\alpha < 0$. (This case was included in [14].) In this case, as is well known, we must have $\alpha = -1/d$ with $d \geqslant 1$ an integer; the random tree $\mathcal{T}_n$ then is a random d-ary recursive tree [15, Section 1.3.3]. (The case $d = 1$ is trivial, and we assume $d \geqslant 2$.) In this case the replacements (14.44) do not satisfy (A5′), since $\alpha + \xi$ may take the non-integer negative value $-1/d$. This is can be remedied as in Remark 8.1 by multiplying the number of white balls by d and adjusting the activity a_1, but in the present case we find it simpler to multiply the number of balls of both colours by d and keep the activities $a_i = 1$. This gives the replacement matrix

$$\begin{pmatrix} d\xi - 1 & d - d\xi \\ 0 & d - 1 \end{pmatrix}, \tag{14.50}$$

with the initial state $\mathbf{x} = (d, 0)$. In (14.45), we have to replace X_{n1} by X_{n1}/d, which yields

$$Z_n = \frac{X_{n1} + N_{n1}}{d}. \tag{14.51}$$

We have $\lambda_1^* = \lambda_1 = \mathbb{E}\,\xi_{11} = dp - 1$, so (A7) requires $dp > 1$. (In fact, it is easily seen that if $dp \leqslant 1$, then $Z_n \xrightarrow{\text{a.s.}} Z_\infty < \infty$ [14].) Assuming $dp > 1$, this urn satisfies (A1)–(A4), (A5′), and (A6)–(A7). It does not satisfy (A8), but Lemma 10.9 shows that $\mathcal{X}_0 > 0$ a.s., and thus we obtain by Theorem 8.5 the a.s. limit (14.49) again (with $\widehat{\mathcal{X}}_1$ replaced by $\widehat{\mathcal{X}}_1/d$); this can be written

$$\frac{Z_n}{n^{(dp-1)/(d-1)}} \xrightarrow{\text{a.s.}} \mathcal{Z} := \frac{p}{dp-1}\widehat{\mathcal{X}}_1. \tag{14.52}$$

Moment convergence, earlier shown by [14], follows from Theorems 12.5 and 12.8. △

Example 14.9. In Example 14.7, we studied the number Z_n of vertices in the preferential attachment tree $\mathcal{T}_n$ such that the path to the root contains only active vertices. More generally, let $Z_n^{(k)}$ be the number of vertices such that this path contains exactly $k \geqslant 0$ passive edges. For fixed k, this can be treated similarly, with an urn with $q \geqslant k + 2$ colours $1, \dots, q$, where vertices with j passive edges on the path to the root are represented by colour $\max(j + 1, q)$. For example, for $q = 3$, this leads to an urn with replacement matrix

$$\begin{pmatrix} \alpha+\xi & 1-\xi & 0 \\ 0 & \alpha+\xi & 1-\xi \\ 0 & 0 & \alpha+1 \end{pmatrix}, \tag{14.53}$$

where $\xi \in \mathrm{Be}(p)$ as above. (It does not matter whether we write this with the same ξ on both rows or not; recall Remark 1.4.) For a general $q \geqslant 2$ we have $\xi_{ii} = \alpha + \xi$ and $\xi_{i,i+1} = 1 - \xi$ for $1 \leqslant i \leqslant q - 1$, $\xi_{qq} = \alpha + 1$, and all other $\xi_{ij} = 0$. The urn is balanced with balance $b = \alpha + 1$. We assume for simplicity $\alpha \geqslant 0$; the case $\alpha < 0$ can be treated as in Example 14.8.

We find $\lambda_1 = \cdots = \lambda_{q-1} = \alpha + p$ and $\lambda_q = \alpha + 1$, and thus $\lambda_1^* = \cdots = \lambda_{q-1}^* = \alpha + p$, $\widehat{\lambda} = \lambda_q^* = \alpha + 1$, $\kappa_i = i - 1$ for $1 \leqslant i \leqslant q - 1$, and $\kappa_q = 0$. We have in analogy with (14.45), $Z_n^{(k)} = X_{n,k+1} - \alpha N_{n,k+1}$ for $k \leqslant q - 2$. Theorem 1.8 applies and shows together with Theorem 11.4 as in Example 14.7, an a.s. limit. Since $\kappa_{k+1} = k$, we now get

$$\frac{Z_n^{(k)}}{n^{(\alpha+p)/(\alpha+1)}\log^k n} \xrightarrow{\text{a.s.}} \mathcal{Z}^{(k)} := \frac{p}{\alpha+p}\widehat{\mathcal{X}}_{k+1}. \tag{14.54}$$

Furthermore, we obtain from (7.3) and induction, since $r_{i,i+1} = \mathbb{E}(1-\xi) = 1-p$ for every $i \leqslant q-1$,

$$\mathcal{Z}^{(k)} = \frac{1}{k!}\left(\frac{1-p}{\alpha+1}\right)^k \mathcal{Z}^{(0)}, \tag{14.55}$$

where $\mathcal{Z}^{(0)}$ equals $\mathcal{Z}$ in (14.49). Consequently, the numbers $Z_n^{(k)}$ for different k are asymptotically proportional. △

Example 14.10. Another example where the urn in Example 14.5 with replacement matrix (14.32) appears is for elephant random walks with delays. In a standard elephant random walk (ERW), the elephant takes steps $Y_n \in \{\pm 1\}$; after an initial step Y_1, the elephant (which remembers the entire walk) chooses one of the preceding steps, uniformly at random, and then randomly either (with probability p) repeats it, or (with probability q) takes a step in the opposite direction. (Here $q = 1-p$.) As noted by Baur and Bertoin [7] (to which we refer for details and further references), this may be modelled by a (non-triangular) Pólya urn, with one white ball for each step $+1$ and one black ball for each step -1 taken so far; the replacement matrix is

$$\begin{pmatrix} \xi & 1-\xi \\ 1-\xi & \xi \end{pmatrix} \tag{14.56}$$

with $\xi \in \text{Be}(p)$. Hence results for the ERW follows from known results for irreducible Pólya urns [7]. (We have nothing to add here.)

In the *elephant random walk with delays* [33; 23; 22], the elephant has a third possibility: with probability r it makes a step 0 (i.e., stays put), regardless of the remembered step. (Now $p+q+r = 1$; we assume $p, q, r > 0$.) This can be modelled by a 3-colour urn, with colours representing steps $+1$, -1, and 0, and replacement matrix

$$\begin{pmatrix} \zeta_1 & \zeta_2 & \zeta_3 \\ \zeta_2 & \zeta_1 & \zeta_3 \\ 0 & 0 & 1 \end{pmatrix}, \tag{14.57}$$

where $(\zeta_1, \zeta_2, \zeta_3)$ is a random vector with exactly one component 1 and the others 0, and $(P[\zeta_i = 1])_{i=1}^3 = (p, q, r)$. This Pólya urn is neither triangular nor irreducible, but it may be regarded as a combination of two such urns (cf. Remark 2.3) as follows. Let as before $Y_n \in \{\pm 1, 0\}$ be the nth step, and let $Z_n := |Y_n| \in \{0, 1\}$; Z_n thus just records whether the elephant moves or stays put. (The process Z_n is called *Bernoulli elephant random walk* in [25]; it has also been studied in [8] and, for somewhat different reasons, in [26].) As noted by [7, V.C], the process (Z_n) can be modelled by a Pólya urn with $q = 2$ and one white ball for each step ± 1 and one black ball for each step 0 so far; the replacement matrix is

$$\begin{pmatrix} 1-\zeta_3 & \zeta_3 \\ 0 & 1 \end{pmatrix}. \tag{14.58}$$

This is the urn in Example 14.5 with $\xi = 1-\zeta_3 \in \text{Be}(1-r)$. The number of non-zero steps up to time n is W_n, the number of white balls in the urn. By conditioning on $Z_1 := |Y_1|$, we may assume that Z_1 is deterministic. Moreover, the case $Z_1 = 0$

is trivial, with $Y_n = Z_n = 0$ for all $n \geqslant 1$; hence we may assume $Z_1 = 1$, and thus the urn starts with 1 white ball. ([22; 24] use a different initial condition with $Z_1 \sim \mathrm{Be}(1-r)$.) Then the urn is the same as in Example 14.6, and Theorem 1.8 yields

$$W_n/n^{1-r} \to \widehat{\mathcal{X}}_1, \tag{14.59}$$

as shown by other methods in [22, Theorem 3.1]; furthermore, [8, Lemma 2.1] and [24, Theorem 5.1] show that (14.43) holds (with p replaced by $1-r$ and r by s, say), and thus $\widehat{\mathcal{X}}_1$ has a Mittag-Leffler distribution (as we saw in Example 14.6). Moreover, [8] and [25] show that moment convergence holds in (14.59); this also follows from Theorem 12.5.

Conditioned on the number W_n, the position of the elephant is the same as for a standard ERW with W_n steps, and thus the limit results in [8] and [24] for the ERW with delays may easily be obtained by combining (14.59) and the results for the standard ERW obtained by [7] from the urn (14.56); we leave the details to the reader. △

Example 14.11. The triangular urn with $q = 3$ and balanced deterministic replacements (with all entries integers $\geqslant 0$)

$$\begin{pmatrix} \alpha & \beta & \sigma-\alpha-\beta \\ 0 & \delta & \sigma-\delta \\ 0 & 0 & \sigma \end{pmatrix} \tag{14.60}$$

was studied by Puyhaubert [43, Section 2.5] and Flajolet, Dumas and Puyhaubert [17, Section 10]; the results include convergence in distribution (after normalization), and, in some cases, convergence of all moments with explicit formulas for moments of the limits.

In particular, they show (in our notation and correcting several typos) that if $\alpha > \delta > 0$ and $\beta > 0$, then

$$X_{n2}/n^{\alpha/\sigma} \xrightarrow{\mathrm{d}} \widehat{\mathcal{X}}_2 \tag{14.61}$$

where the limit has moments

$$\mathbb{E}\,\widehat{\mathcal{X}}_2^r = \left(\frac{\alpha\beta}{\alpha-\delta}\right)^r \frac{\Gamma(x_1/\alpha + r)\Gamma(|\mathbf{x}|/\sigma)}{\Gamma(x_1/\alpha)\Gamma((|\mathbf{x}|+r\alpha)/\sigma)}. \tag{14.62}$$

In this case, we have $\sigma \geqslant \alpha+\beta > \alpha > \delta$ and thus $\lambda_1^* = \lambda_2^* = \lambda_1 = \alpha$, $\lambda_2 = \delta$, $\widehat{\lambda} = \lambda_3^* = \lambda_3 = \sigma$, $\widehat{\kappa} = \kappa_i = 0$, $\gamma_i = 0$ $(i = 1,2,3)$. Hence, Theorem 1.8 yields (14.61) with the stronger convergence a.s. Moreover, the leaders are 1 and 3, and Lemma 7.3 yields $\widehat{\mathcal{X}}_2 = \widehat{c}_{21}\widehat{\mathcal{X}}_1$, where Lemma 7.3 and (6.18) show that $(1, \widehat{c}_{21})$ is a left eigenvector of $\left(\begin{smallmatrix} \alpha & \beta \\ 0 & \delta \end{smallmatrix}\right)$, with eigenvalue α. Consequently, $\widehat{c}_{21} = \beta/(\alpha-\delta)$ and

$$\widehat{\mathcal{X}}_2 = \frac{\beta}{\alpha-\delta}\widehat{\mathcal{X}}_1. \tag{14.63}$$

Furthermore, as noted in [17], we may in this urn be partially colour-blind and merge colours 2 and 3; then $(X_{n1}, X_{n2}+X_{n3})$ is a 2-colour urn with replacement matrix $\left(\begin{smallmatrix} \alpha & \sigma-\alpha \\ 0 & \sigma \end{smallmatrix}\right)$; hence the moments $\mathbb{E}\,\widehat{\mathcal{X}}_1^r$ are given by (14.30), where now α and δ are replaced by σ and α, and $x_1 + x_2$ is replaced by $|\mathbf{x}| = x_1 + x_2 + x_3$, i.e.,

$$\mathbb{E}\,\widehat{\mathcal{X}}_1^r = \alpha^r \frac{\Gamma\bigl(|\mathbf{x}|/\sigma\bigr)\Gamma(x_1/\alpha + r)}{\Gamma(x_1/\alpha)\Gamma\bigl((|\mathbf{x}|+r\alpha)/\sigma\bigr)}, \qquad r \geqslant 0. \tag{14.64}$$

Thus, (14.62) follows by (14.63).

Similarly, in the case $\alpha = \delta > 0$ and $\beta > 0$, [43] and [17] show (again correcting several typos)

$$X_{n2}/(n^{\alpha/\sigma}\log n) \xrightarrow{\mathrm{d}} \widehat{\mathcal{X}}_2 \tag{14.65}$$

where the limit has moments

$$\mathbb{E}\,\widehat{\mathcal{X}}_2^r = \Bigl(\frac{\alpha\beta}{\sigma}\Bigr)^r \frac{\Gamma(x_1/\alpha + r)\Gamma(|\mathbf{x}|/\sigma)}{\Gamma(x_1/\alpha)\Gamma((|\mathbf{x}| + r\alpha)/\sigma)}. \tag{14.66}$$

In this case $\lambda_1^* = \lambda_2^* = \lambda_1 = \lambda_2 = \alpha$, $\kappa_1 = 0$ and $\kappa_2 = 1$, and as above $\widehat{\lambda} = \lambda_3^* = \lambda_3 = \sigma$ and $\widehat{\kappa} = \kappa_3 = 0$. Hence, Theorem 1.8 yields (14.65) with convergence a.s. Moreover, Lemma 7.3 and (7.3) yield

$$\widehat{\mathcal{X}}_2 = \frac{\beta}{\sigma}\widehat{\mathcal{X}}_1, \tag{14.67}$$

which yields (14.66) by (14.64).

[17] and [43] further prove moment convergence in (14.61) and (14.65), which also follows from Theorem 12.5.

Note that [17] and [43] assume the replacements to be integers, while we obtain the results above also for non-integer replacements.

A.s. convergence in (14.61) and (14.65) follows also from [12], see Example 14.12. △

Example 14.12. Bose, Dasgupta, and Maulik [12] study a rather general class of balanced triangular urns with deterministic $\boldsymbol{\xi}_i$; thus $\xi_{ij} = r_{ij}$. They show a.s. convergence of X_{ni} suitably normalized, as in our Theorem 1.8.

We introduce some of the notation from [12]. The colours are $\{1, \dots, q\}$ (where they write $q = K + 1$), and the replacement matrix is triangular, so (2.4)–(2.6) hold with the natural order $<$. The diagonal entries $r_{ii} = \lambda_{ii}$ are denoted r_i, and the positions of the weak maxima in the sequence $r_1, \dots, r_q$ are denoted $i_1, \dots, i_{J+1}$. Thus $r_{i_1} \leqslant r_{i_2} \leqslant \dots \leqslant r_{i_{J+1}}$, and $r_k < r_{i_j}$ when $i_j < k < i_{j+1}$. Since the urn is balanced, Lemma 10.7 and Remark 10.8 show that $r_q = \lambda_q = \widehat{\lambda}$ is a maximum, and thus $i_{J+1} = q$. Clearly, $i_1 = 1$. The *jth block* of colours is $\{i_j, \dots, i_{j+1} - 1\}$. (In [12], the replacements are normalized by $\lambda_q = \widehat{\lambda} = 1$, which can be assumed without loss of generality. It is also assumed that initially there is 1 ball in the urn, i.e., $\sum_i x_i = 1$; this seems to be a mistake since one cannot in general normalize both to 1 simultaneously.)

[12] says that the colours are arranged in *increasing order* if for every $k \in (i_j, i_{j+1})$ (with $1 \leqslant j \leqslant J$), there exists $m \in [i_j, k)$ such that $r_{mk} > 0$. Using our terminology, this is easily seen to be equivalent to: If $k \in (i_j, i_{j+1})$, then k is a descendant of i_j. [12, Proposition 2.1] shows that in every balanced triangular urn, the colours can be rearranged in increasing order. [Sketch of proof: Construct the blocks in backwards order. In each step find a colour i (to be labelled i_j) with λ_i maximal among the remaining colours; let the next block consist of i and all its remaining descendants. Order this block in a suitable way, with i first, and place it before the previously constructed blocks. Repeat with the remaining colours.]

The main result of [12] further assumes [12, (2.2)], which says that for every $j = 1, \dots, J$, there exists $m \in [i_j, i_{j+1})$ such that $r_{m,i_{j+1}} > 0$. In our terminology, and assuming (as in [12]) that the colours are in natural order, this is equivalent to $i_{j+1} \succ i_j$. It follows that the assumptions in [12] (i.e., increasing order and their (2.2)) imply that $i_1, \dots, i_{J+1}$ are precisely the colours i with $\lambda_i^* = \lambda_i$, i.e. our leaders and subleaders (see (6.3) and (6.19)), and that the leaders ν have distinct λ_ν^*. Moreover, for each leader $\nu = i_j$, the subleaders in D_ν (see (6.13)) are $i_{j+1}, \dots, i_{j+\ell}$ where $\ell = \ell_j \geqslant 0$ is the largest integer with $\lambda_{i_{j+\ell}} = \lambda_{i_j}$; these subleaders form a chain in the partial order $\prec$, and thus $\kappa_{i_{j+k}} = k$ for $k = 0, \dots, \ell$. The set D_ν^κ is an interval $[i_j, i_{j+1})$ from a (sub)leader to the next.

The a.s. convergence in the main result [12, Theorem 3.1] now follows from Theorem 1.8, Lemma 6.7, and Theorem 7.1. Moreover, [12, Remark 3.3] says that it is clear from their proof that a.s. convergence holds also without their assumptions of increasing order and their (2.2); [12] thus essentially states our Theorem 1.8 for the case of a balanced urn with deterministic replacements.

Furthermore, [12] also shows convergence in L^2, which is extended to L^p for any p by our Theorem 12.5. △

Example 14.13. Consider an urn with $q = 4$ and replacement matrix

$$\begin{pmatrix} \alpha\eta_1 - 1 & 0 & \eta_2 & 0 \\ 0 & \beta\eta_1 - 1 & \eta_2 & 0 \\ 0 & 0 & \gamma & 0 \\ 0 & 0 & 0 & \delta \end{pmatrix}, \tag{14.68}$$

where $\alpha, \beta, \gamma, \delta$ are positive integers and $\eta_1, \eta_2 \in \mathrm{Be}(1/2)$ are independent. We have $\lambda_1 = \alpha/2 - 1$, $\lambda_2 = \beta/2 - 1$, $\lambda_3 = \gamma$, and $\lambda_4 = \delta$. Suppose that $\lambda_4 = \lambda_1 > \lambda_2 > \lambda_3 > 0$, and start with $\mathbf{X}_0 = (1, 1, 0, 1)$.

This urn satisfies (A1)–(A4), (A5′), and (A6)–(A7), but not (A8). We have $\widehat{\lambda} = \lambda_1^* = \lambda_3^* = \lambda_4^* = \lambda_1$, $\lambda_2^* = \lambda_2$, and $\kappa_i = 0\ \forall i$. Theorem 8.6 shows that (4.1) holds for all colours i, with

$$e^{-\lambda_i^* t} X_i(t) \xrightarrow{\text{a.s.}} \mathcal{X}_i. \tag{14.69}$$

Furthermore, $\mathcal{X}_4 > 0$ a.s., as is seen by considering only balls of colour 4; hence, by (5.17), $\mathcal{X}_0 = \widehat{\lambda}^{-1}(\mathcal{X}_1 + \mathcal{X}_4) \geqslant \widehat{\lambda}^{-1}\mathcal{X}_4 > 0$ a.s.

Furthermore, 3 is a follower of the leader 1, and Lemma 6.1 with Remark 8.8 yields $\mathcal{X}_3 = c\mathcal{X}_1$ for some $c > 0$. (In fact, $c = 1/2(\lambda_1 - \lambda_3)$, by the same argument as for (14.13).) Note also that $\mathcal{X}_1$ and $\mathcal{X}_2$ are independent.

It is obvious that colours 1 and 2 both may die out in a few draws, and that if they do, they may or may not first generate a ball of colour 3. If they do not die out, then a.s. $\mathcal{X}_1 > 0$ and $\mathcal{X}_2 > 0$ respectively, see Remark 8.10.

Consequently, the following cases can appear, all with positive probabilities:

(i) $\mathcal{X}_1 > 0$ and then $\mathcal{X}_3 > 0$ and thus $X_3(t)$ grows at rate $e^{\lambda_3^* t} = e^{\lambda_1 t}$. Similarly, by Theorem 8.5 and (5.12), $X_{n3}/n \xrightarrow{\text{a.s.}} \widehat{\mathcal{X}}_3 > 0$.

(ii) $\mathcal{X}_1 = 0$ but $\mathcal{X}_2 > 0$; then $\mathcal{X}_3 = 0$, but by considering the urn with colour 2, 3, and 4 only (after 1 has died out), it follows that $e^{-\lambda_2} X_3(t) \xrightarrow{\text{a.s.}} \mathcal{X}_3' := c'\mathcal{X}_2 > 0$ for some $c' > 0$, and thus $X_3(t)$ grows at rate $e^{\lambda_2 t}$; similarly $X_{n3}/n^{\lambda_2/\lambda_1} \xrightarrow{\text{a.s.}} \widehat{\mathcal{X}}_3' > 0$.

(iii) $\mathcal{X}_1 = \mathcal{X}_2 = 0$ and both $X_1(t)$ and $X_2(t)$ die out, but at least one of them first gets a ball of colour 3 as offspring. Then, by considering only balls of colour 3 and 4 (when the others have died out), $e^{-\lambda_3} X_3(t) \xrightarrow{\text{a.s.}} \mathcal{X}_3'' > 0$, and thus $X_3(t)$ grows at rate $e^{\lambda_3 t}$; similarly $X_{n3}/n^{\lambda_3/\lambda_1} \xrightarrow{\text{a.s.}} \widehat{\mathcal{X}}_3'' > 0$.

(iv) $\mathcal{X}_1 = \mathcal{X}_2 = 0$, and both $X_1(t)$ and $X_2(t)$ die out without producing an offspring of colour 3. Then, $X_3(t) = 0$ for all t and thus $X_{3n} = 0$ for all n.

This example shows that in a case when $\mathcal{X}_i = 0$ with positive probability, it may still be possible to find precise limit results, but different limit results may hold in different subcases. It seems that this can be done very generally on a case by case basis, but as said in Remark 8.7 we do not attempt any general statement. △

Example 14.14. An interesting counterexample is given by the replacement matrix

$$\begin{pmatrix} 0 & 1 \\ 0 & \pm 1 \end{pmatrix}, \tag{14.70}$$

where $\xi_{22} = \pm 1$ denotes a random variable with $\mathbb{P}(\xi_{22} = 1) = \mathbb{P}(\xi_{22} = -1) = \frac{1}{2}$. In words: when we draw a white ball, it is replaced together with a black ball; when we draw a black ball, we toss a coin and then (with probability $\frac{1}{2}$ each) either remove the ball or replace it together with another black ball.

We have $r_{22} = \mathbb{E}\,\xi_{22} = 0$, and thus $\lambda_1 = \lambda_2 = 0 = \lambda_1^* = \lambda_2^* = \widehat{\lambda}$, $\kappa_1 = 0$, $\kappa_2 = 1$, $\widehat{\kappa} = 1$, $\widehat{\kappa}_0 = 2$. Note that this example is excluded from Theorems 8.4–8.6 since (A7) does not hold. (However, (A1)–(A4), (A5′), and (A6) hold, provided $x_1 > 0$ and $x_2 \in \mathbb{Z}_{\geqslant 0}$.) We will see that, in fact, we do *not* have a.s. convergence of $B_n/n^{1/2} = X_{n2}/n^{1/2}$ and $B(t)/t = X_2(t)/t$ as (1.4) and (4.1) would give; however, these converge in distribution.

Consider first a continuous time urn with only black balls. This is a branching process where balls live an Exp(1) time and then randomly either die or are split into two. Let $Y(t)$ denote such an urn with black balls, starting with $Y(0) = 1$, and denote its probability generating function by (for $|z| \leqslant 1$, say)

$$g_t(z) := \mathbb{E}\, z^{Y(t)}. \tag{14.71}$$

Then the Kolmogorov backward equation [32, Theorem 12.22] yields, for $t > 0$,

$$\frac{\partial}{\partial t} g_t(z) = \frac{1}{2}\bigl(1 + g_t(z)^2\bigr) - g_t(z) = \frac{1}{2}\bigl(1 - g_t(z)\bigr)^2 \tag{14.72}$$

with $g_0(z) = z$, which has the solution

$$g_t(z) = 1 - \Bigl(\frac{1}{1-z} + \frac{t}{2}\Bigr)^{-1} = \frac{t(1-z) + 2z}{t(1-z) + 2} = \frac{t}{t+2} + \frac{2}{t+2} \cdot \frac{\frac{2}{t+2} z}{1 - \frac{t}{t+2} z}. \tag{14.73}$$

Consequently, $Y(t)$ has a modified geometric distribution: $\mathbb{P}(Y(t) = 0) = g_t(0) = \frac{t}{t+2}$ and the conditional distribution $\bigl(Y(t) \mid Y(t) > 0\bigr)$ is $\mathrm{Ge}\bigl(\frac{2}{t+2}\bigr)$. In particular, $\mathbb{P}(Y(t) = 0) \to 1$ as $t \to \infty$; since 0 is an absorbing state, it follows that a.s. $Y(t) = 0$ for sufficiently large t; in other words, $Y(t)$ dies out. (This follows also since $Y(t)$ is a time-changed simple random walk, absorbed at 0.) A simple calculation yields $\mathbb{E}\, Y(t) = 1$ and $\operatorname{Var} Y(t) = t$, in accordance with (8.4)–(8.5) and (8.11). Note that $\bigl(t^{-1} Y(t) \mid Y(t) > 0\bigr) \xrightarrow{\mathrm{d}} \mathrm{Exp}\bigl(\frac{1}{2}\bigr)$ as $t \to \infty$. Roughly speaking, for large t, $Y(t)$ is non-zero with probability $\approx 2/t$, and if it is, it is of order t.

Now consider the two-colour urn above, and assume that we start with $\mathbf{X}_0 = (1, 0)$, i.e., 1 white ball. The number of white balls is constant for this urn, and thus $W(t) = 1$ for all t in the continuous-time urn. This means that white balls are drawn according to a Poisson process Ξ with constant rate 1. If the times they are drawn are $(T_k)_1^\infty$, then we have

$$B(t)=\sum_{T_k\leqslant t} Y_k(t-T_k), \tag{14.74}$$

where Y_k are independent copies of the one-colour process $Y(t)$, independent also of $(T_k)_1^\infty$. Consequently, for $z\in[0,1]$ say, we have

$$\begin{aligned}\mathbb{E}\left(z^{B(t)}\mid\Xi\right)&=\prod_{T_k\leqslant t} g_{t-T_k}(z)=\exp\Bigl(\sum_{T_k\leqslant t}\log g_{t-T_k}(z)\Bigr)\\&=\exp\Bigl(\int_0^t \log g_{t-u}(z)\,\mathrm{d}\Xi(u)\Bigr).\end{aligned} \tag{14.75}$$

Hence, by a standard formula for Poisson processes [32, Lemma 12.2],

$$\begin{aligned}\mathbb{E}\,z^{B(t)}&=\mathbb{E}\,\exp\Bigl(\int_0^t \log g_{t-u}(z)\,\mathrm{d}\Xi(u)\Bigr)=\exp\Bigl(\int_0^t \bigl(g_s(z)-1\bigr)\,\mathrm{d}s\Bigr)\\&=\exp\Bigl(-\int_0^t\Bigl(\frac{1}{1-z}+\frac{s}{2}\Bigr)^{-1}\mathrm{d}s\Bigr)=\exp\Bigl(-2\Bigl(\log\Bigl(\frac{1}{1-z}+\frac{t}{2}\Bigr)-\log\Bigl(\frac{1}{1-z}\Bigr)\Bigr)\Bigr)\\&=\Bigl(1+\frac{t}{2}(1-z)\Bigr)^{-2}=\Bigl(\frac{\frac{2}{t+2}}{1-\frac{t}{t+2}z}\Bigr)^2.\end{aligned} \tag{14.76}$$

Consequently, $B(t)$ has a negative binomial distribution $\mathrm{NegBin}\bigl(2,\frac{2}{t+2}\bigr)$. In particular,

$$\mathbb{E}\,B(t)=t, \tag{14.77}$$

which also follows directly from (14.74).

It follows from (14.76) that as $t\to\infty$, for any $s>0$,

$$\mathbb{E}\,e^{-sB(t)/t}=\Bigl(1+\frac{t}{2}(1-e^{-s/t})\Bigr)^{-2}\to\Bigl(1+\frac{s}{2}\Bigr)^{-2} \tag{14.78}$$

and thus $B(t)/t$ converges in distribution to a Gamma distribution:

$$t^{-1}B(t)\overset{\mathrm{d}}{\longrightarrow}\Gamma(2,\tfrac12),\qquad\text{as } t\to\infty. \tag{14.79}$$

However, we will see that $B(t)/t$ does *not* converge a.s., which shows that Theorem 8.6 does not extend to this example.

To see this, we extend (14.79) to process convergence. We claim that as $t\to\infty$, we have

$$t^{-1}B(tx)\overset{\mathrm{d}}{\longrightarrow}\mathcal{B}(x):=\tfrac14\mathrm{BESQ}^4(x)\qquad\text{in } D[0,\infty), \tag{14.80}$$

where $\mathrm{BESQ}^4(x)$ denotes a squared 4-dimensional Bessel process [44, Chapter XI]. Recall that

$$\mathrm{BESQ}^4(x)=|\mathcal{W}(x)|^2=\sum_{i=1}^4\mathcal{W}_i(x)^2, \tag{14.81}$$

where $\mathcal{W}_1(x),\dots,\mathcal{W}_4(x)$ are independent standard Brownian motions (Wiener processes), and $\mathcal{W}(x):=\bigl(\mathcal{W}_1(x),\dots,\mathcal{W}_4(x)\bigr)$ thus is a 4-dimensional Brownian motion. Hence, $\frac14\mathrm{BESQ}^4(x)\sim\Gamma(2,x/2)$, in accordance with (14.79). Furthermore, a.s. $\mathrm{BESQ}^4(x)>0$ for every $x>0$.

The proof of (14.80) is somewhat technical and is given in Appendix C, where we also extend the result to other initial values (w_0,b_0), see Theorem C.1.

Suppose now that $B(t)/t \xrightarrow{\text{a.s.}} \mathcal{Z}$ for some random variable $\mathcal{Z}$; then $\mathcal{Z} \sim \Gamma(2, \frac{1}{2})$ by (14.79). Moreover, for every $r > 0$, we would have

$$t^{-1}\bigl(B(rt) - rB(t)\bigr) = r\Bigl(\frac{B(rt)}{rt} - \frac{B(t)}{t}\Bigr) \xrightarrow{\text{a.s.}} r(\mathcal{Z} - \mathcal{Z}) = 0. \tag{14.82}$$

However, the process convergence in (14.80) implies finite dimension convergence, and in particular

$$t^{-1}\bigl(B(rt) - rB(t)\bigr) \xrightarrow{\text{d}} \frac{1}{4}\bigl(\mathrm{BESQ}^4(r) - r\mathrm{BESQ}^4(1)\bigr). \tag{14.83}$$

Hence, (14.82) would imply $\mathrm{BESQ}^4(r) = r\mathrm{BESQ}^4(1)$ a.s., for every $r > 0$, which obviously is false (even for a single $r \neq 1$). This contradiction proves the claim that $B(t)/t$ does not converge a.s. Hence, the convergence in (14.79) holds in distribution but not a.s.

We use (14.80) to derive corresponding results for the discrete-time urn. Let, as above, $\widehat{T}_n$ be the nth time that a ball is drawn, and let $N(t)$ be the total number of draws up to time t; thus $N(\widehat{T}_n) = n$. Since all balls are drawn with intensity 1,

$$\widetilde{N}(t) := N(t) - \int_0^t \bigl(W(s) + B(s)\bigr)\,\mathrm{d}s = N(t) - t - \int_0^t B(s)\,\mathrm{d}s \tag{14.84}$$

is a local martingale with $\widetilde{N}(0) = 0$, and it follows as in the proof of Lemma 3.1 (now using (14.77)) that $\widetilde{N}(t)$ is a martingale. In particular $\mathbb{E}\,\widetilde{N}(t) = 0$, and thus by (14.84) and (14.77)

$$\mathbb{E}\,N(t) = t + \mathbb{E}\int_0^t B(s)\,\mathrm{d}s = t + \int_0^t \mathbb{E}\,B(s)\,\mathrm{d}s = t + t^2/2. \tag{14.85}$$

Furthermore, all jumps are $+1$ and thus the quadratic variation is by (2.16)

$$[\widetilde{N}, \widetilde{N}]_t = \sum_{0<s\leqslant t} \Delta N(s) = N(t). \tag{14.86}$$

Consequently, by Doob's inequality (2.18) and (14.85),

$$\mathbb{E}\,\widetilde{N}^*(t)^2 \leqslant C\,\mathbb{E}\,[\widetilde{N}, \widetilde{N}]_t = C\,\mathbb{E}\,N(t) = Ct + Ct^2. \tag{14.87}$$

In particular, $\widetilde{N}^*(t)/t^2 \xrightarrow{\text{p}} 0$ as $t \to \infty$, which together with (14.84) implies that

$$N(tx)/t^2 - \int_0^x t^{-1}B(ty)\,\mathrm{d}y = t^{-2}N(tx) - t^{-2}\int_0^{tx} B(s)\,\mathrm{d}s = t^{-2}\widetilde{N}(tx) + t^{-1}x \xrightarrow{\text{p}} 0 \tag{14.88}$$

in $D[0, \infty)$, and consequently (14.80) implies

$$N(tx)/t^2 \xrightarrow{\text{d}} \mathcal{V}(x) := \int_0^x \mathcal{B}(y)\,\mathrm{d}y, \tag{14.89}$$

in $D[0, \infty)$, jointly with (14.80). Note that $\mathcal{V}(x)$ is a continuous stochastic process which strictly increases from $\mathcal{V}(0) = 0$ to ∞. Define τ by

$$\mathcal{V}(\tau) = 1; \tag{14.90}$$

thus τ is random with $0 < \tau < \infty$ a.s. It follows easily from (14.89) (we omit the details) that, jointly with (14.80),

$$\widehat{T}_n/\sqrt{n} \xrightarrow{\text{p}} \tau \tag{14.91}$$

and as a consequence

$$B_n/\sqrt{n} = B(\widehat{T}_n)/\sqrt{n} \xrightarrow{\mathrm{d}} \mathcal{B}(\tau). \tag{14.92}$$

This proves convergence in distribution of $B_n/\sqrt{n}$. The limit $\mathcal{B}(\tau)$ is determined by (14.80), (14.89), and (14.90); unfortunately we do not know any simpler description of the limit distribution, and we leave it as an open problem to find one.

For the discrete-time urn, we thus have convergence in distribution of $B_n/n^{1/2}$. (The exponent 1/2 equals $\kappa_2/\widehat{\kappa}_0$, just as in Theorem 1.8(ii) although we cannot apply that theorem.) However, we do not have convergence a.s. in (14.92). In fact, if $B_n/n^{1/2} \xrightarrow{\text{a.s.}} \widehat{\mathcal{Z}}$ for some $\widehat{\mathcal{Z}}$, then for every $x > 0$, as $t \to \infty$,

$$\frac{B(xt)^2}{N(xt)} = \frac{B^2_{N(xt)}}{N(xt)} \xrightarrow{\text{a.s.}} \widehat{\mathcal{Z}}^2, \tag{14.93}$$

while the joint convergence of (14.80) and (14.89) implies

$$\frac{B(xt)^2}{N(xt)} = \left(\frac{B(xt)}{t}\right)^2 \cdot \frac{t^2}{N(xt)} \xrightarrow{\mathrm{d}} \frac{\mathcal{B}(x)^2}{\mathcal{V}(x)} \tag{14.94}$$

in $D(0,\infty)$, i.e., in $D[a,b]$ for every $0 < a < b < \infty$. Consequently, (14.93) would imply that a.s. $\mathcal{B}(x)^2/\mathcal{V}(x) = \widehat{\mathcal{Z}}^2$ for every $x > 0$, and thus a.s.

$$\int_0^x \mathcal{B}(y)\,\mathrm{d}y = \mathcal{V}(x) = \widehat{\mathcal{Z}}^{-2}\mathcal{B}(x)^2, \qquad x > 0. \tag{14.95}$$

But this impossible, for example because (14.95) would imply that $\mathcal{B}(x)$, and thus $\mathrm{BESQ}^4(x)$, is differentiable (and, moreover, linear). This contradiction shows that $B_n/\sqrt{n}$ does not converge a.s. △

Example 14.15. A counterexample somewhat similar to Example 14.14 is given by the replacement matrix

$$\begin{pmatrix} 0 & 1 \\ 0 & -1 \end{pmatrix}. \tag{14.96}$$

In words: when we draw a white ball, it is replaced together with a black ball; when we draw a black ball, we discard it.

We have $\lambda_1 = 0$ and $\lambda_2 = -1$, and thus $\lambda_1^* = \lambda_2^* = 0$. Note that this example too is excluded from Theorem 8.6 since (A7) does not hold. (However, again (A1)–(A4), (A5′), and (A6) hold, provided $x_1 > 0$ and $x_2 \in \mathbb{Z}_{\geqslant 0}$.) We will see that, as in Example 14.14, $B(t)$ converges in distribution but *not* a.s.

Suppose that the urn starts with a single white ball, i.e., $\mathbf{X}_0 = (1,0)$. Then, for the continuous-time urn, $W(t) = 1$ for all t, and thus white balls are drawn according to a Poisson process with constant rate 1. At each draw in this Poisson process, we add a black ball. Black balls live an exponential time with mean 1, and then disappear. Consequently, $B(t)$, the number of black balls, is a birth-death process where the birth rate is constant 1 and the death rate equals the number of particles. (Thus $\mu_k = 1$ and $\lambda_k = k$ in the standard notation.)

Let T_k be the time of the kth white draw, and L_k the life-length of the black ball that then is added. Then the pairs (T_k, L_k) form a Poisson process in $\mathbb{R}^2_+$ with rate $e^{-y}\,\mathrm{d}x\,\mathrm{d}y$. The number of black balls at time t is

$$B(t) = \sum_{k=1}^{\infty} \mathbf{1}_{\{T_k \leqslant t\}} \mathbf{1}_{\{L_k > t - T_k\}} = \sum_{k=1}^{\infty} \mathbf{1}_{\{(T_k, L_k) \in D_t\}} \tag{14.97}$$

where $D_t := \{(x,y) \in \mathbb{R}^2 : 0 < x \leqslant t,\, y > t - x\}$. Consequently, $B(t)$ has a Poisson distribution $\mathrm{Po}(\mu(t))$, where

$$\mu(t) = \int_{D_t} e^{-y}\,\mathrm{d}x\,\mathrm{d}y = \int_0^t \int_{t-x}^{\infty} e^{-y}\,\mathrm{d}y\,\mathrm{d}x = \int_0^t e^{x-t}\,\mathrm{d}x = 1 - e^{-t}. \tag{14.98}$$

As $t \to \infty$, we thus have convergence in distribution $B(t) \xrightarrow{\mathrm{d}} \mathrm{Po}(1)$. Obviously, we cannot have convergence a.s., since $B(t)$ jumps $+1$ at every T_k, and $T_k \to \infty$ as $k \to \infty$.

B_n does not converge in distribution as $n \to \infty$ for a simple parity reason: since B_n changes by ± 1 at each draw, we have $B_n \equiv n \pmod 2$. However, B_n is a Markov chain (since W_n is constant), it is irreducible, and it is easily seen that the expected time to return to 0 is finite; hence the Markov chain B_n is positive recurrent. The period is 2, and and it follows that the two subsequences B_{2n} and B_{2n+1} converge in distribution to some limits, say $\widehat{B}_{\mathrm{even}}$ and $\widehat{B}_{\mathrm{odd}}$. The mixture (with equal weights) of $\widehat{B}_{\mathrm{even}}$ and $\widehat{B}_{\mathrm{odd}}$ has a stationary distribution for the Markov chain, and it is easily found that the limit distributions are given by

$$\mathbb{P}(\widehat{B}_{\mathrm{even}} = k) = \frac{k+1}{k!} e^{-1} \mathbf{1}_{\{k \text{ is even}\}}, \tag{14.99}$$

$$\mathbb{P}(\widehat{B}_{\mathrm{odd}} = k) = \frac{k+1}{k!} e^{-1} \mathbf{1}_{\{k \text{ is odd}\}}. \tag{14.100}$$

These can be described as $\mathrm{Po}(1)$ rounded up to nearest even or odd integer, respectively.

Obviously, we do not have convergence a.s., even for these subsequences. △

Example 14.16. Consider a diagonal urn with $q = 2$ and replacement matrix $\begin{pmatrix} \xi_{11} & 0 \\ 0 & \xi_{22} \end{pmatrix}$ as in [2]; cf. the deterministic case in Example 14.2. Assume that $\xi_{11}, \xi_{22} \in \mathbb{Z}_{\geqslant 0}$ a.s., that $\mathbb{E}\,\xi_{11} = \mathbb{E}\,\xi_{22} = 1$, and that $\mathbb{E}\,\xi_{22}^2 < \infty$ but $\mathbb{E}\,\xi_{11} \log \xi_{11} = \infty$, so that (A4) does *not* hold. (We may simply take $\xi_{22} = 1$ a.s.) Then the stochastic processes $X_1(t)$ and $X_2(t)$ are independent Markov branching processes, and [4, Theorem III.7.2] shows that, as $t \to \infty$,

$$e^{-t} X_1(t) \xrightarrow{\mathrm{a.s.}} 0. \tag{14.101}$$

Hence, (4.1) holds with $\mathcal{X}_1 = 0$ a.s., while Theorem 4.1 (applied to colour 2 only), or [4, Theorem III.7.2] again, shows that (4.1) holds also for $i = 2$ with $\mathcal{X}_2 > 0$ a.s. It follows that $X_1(t)/X_2(t) \xrightarrow{\mathrm{a.s.}} 0$ as $t \to \infty$, and thus $X_{n1}/X_{n2} \xrightarrow{\mathrm{a.s.}} 0$ as $n \to \infty$. It follows easily that (1.3) holds, which in this case is $X_{ni}/n \xrightarrow{\mathrm{a.s.}} \widehat{\mathcal{X}}_i$, with $\widehat{\mathcal{X}}_1 = 0$ and $\widehat{\mathcal{X}}_2 = 1$; we thus have $\mathcal{X}_1 = \widehat{\mathcal{X}}_1 = 0$, in contrast to Theorems 1.8 and 4.1.

This shows that the main results in the present paper do not hold without assuming at least $\mathbb{E}\,\xi_{ij} \log \xi_{ij} < \infty$. We have for convenience assumed the stronger second moment condition (A4), but as said in Remark 2.1, we conjecture that it can be weakened. △

Acknowledgement. I thank Allan Gut for help with references.

Appendix A. Absolute continuity and conditioning

In this appendix we state three general lemmas on absolute continuity of distributions and conditioning. We find them intuitively almost obvious, but only almost, and since we do not know any references, we provide complete proofs. For two measures μ and λ on the same space, we let $\mu \ll \lambda$ denote that μ is absolutely continuous with respect to λ, i.e., that $\lambda(B) = 0 \implies \mu(B) = 0$.

We recall some further standard definitions:

A measure space $(\mathfrak{X}, \mathcal{X})$ is a *Borel space* if it is (or is isomorphic to) a Borel set in a complete separable metric space with its Borel σ-field, see [32, Appendix A1]. This includes, for example, $\mathbb{R}$, $\mathbb{R}^n$, and the function space $D[0,\infty)$; moreover, any finite or countable product of Borel spaces is a Borel space.

If $(\mathfrak{X}, \mathcal{X})$ and $(\mathfrak{Y}, \mathcal{Y})$ are two measurable spaces, then a *probability kernel* from $\mathfrak{X}$ to $\mathfrak{Y}$ is a mapping $\mu : \mathfrak{X} \times \mathcal{Y} \to [0,1]$ such that $B \mapsto \mu(x,B)$ is a probability measure on $(\mathfrak{Y}, \mathcal{Y})$ for every fixed $x \in \mathfrak{X}$, and furthermore $x \mapsto \mu(x,B)$ is measurable on $(\mathfrak{X}, \mathcal{X})$ for every fixed $B \in \mathcal{Y}$, see [32, p. 20].

If X and Y are random variables with values in measurable spaces $(\mathfrak{X}, \mathcal{X})$ and $(\mathfrak{Y}, \mathcal{Y})$, respectively, then a *regular conditional distribution* of X, given Y, is a probability kernel μ from $\mathfrak{Y}$ to $\mathfrak{X}$ such that for any fixed $B \in \mathcal{X}$,

$$\mu(Y,B) = \mathbb{P}[X \in B \mid Y] \qquad \text{a.s.,} \tag{A.1}$$

see [32, p. 106–107]. It follows that for any measurable $f : \mathcal{X} \to [0,\infty]$,

$$\mathbb{E}\left[f(X) \mid Y\right] = \int_{\mathfrak{X}} f(x)\mu(Y, \mathrm{d}x) \qquad \text{a.s.} \tag{A.2}$$

Similarly [32, (7) on p. 108], for any measurable $f : \mathcal{X} \times \mathcal{Y} \to [0,\infty]$,

$$\mathbb{E}\, f(X,Y) = \mathbb{E} \int_{\mathfrak{X}} f(x,Y)\mu(Y, \mathrm{d}x). \tag{A.3}$$

If $(\mathfrak{X}, \mathcal{X})$ is a Borel space, then a such a regular conditional distribution μ exists, and the probability measure $\mu(y,\cdot)$ is $\mathcal{L}(Y)$-a.e. unique (in the standard sense that two different such kernels are equal for $\mathcal{L}(Y)$-a.e. $y \in \mathcal{Y}$) [32, Theorem 6.3].

Lemma A.1. *Let X, Y, Z be random variables taking values in Borel spaces $(\mathfrak{X}, \mathcal{X})$, $(\mathfrak{Y}, \mathcal{Y})$, $(\mathfrak{Z}, \mathcal{Z})$, respectively, and suppose that $Z = \varphi(Y)$ for some measurable function $\varphi : \mathfrak{Y} \to \mathfrak{Z}$. Let λ be a measure on $\mathfrak{X}$, and suppose that the regular conditional distribution $\mu(y,\cdot)$ of X given Y is absolutely continuous with respect to λ for $\mathcal{L}(Y)$-a.e. $y \in \mathfrak{Y}$. Then the regular conditional distribution $\mu'(z,\cdot)$ of X given Z is absolutely continuous with respect to λ for $\mathcal{L}(Z)$-a.e. $z \in \mathfrak{Z}$.*

Remark A.2. Although Lemma A.1 is stated for conditionings on single random variables Y and Z, it holds also for conditionings on finite or countably infinite sequences of random variables (taking values in possibly different Borel spaces), since such sequences can be regarded as a single variable in a suitable product space. △

Proof. If $B \subset \mathfrak{X}$ is any set with $\lambda(B) = 0$, then $\mu(Y,B) = 0$ a.s., and thus, by (A.1),

$$\mu'(Z,B) = \mathbb{E}\left[\mathbf{1}_B(X) \mid Z\right] = \mathbb{E}\big[\mathbb{E}\left[\mathbf{1}_B(X) \mid Y\right] \mid Z\big] = \mathbb{E}\left[\mu(Y,B) \mid Z\right] = 0 \quad \text{a.s.} \tag{A.4}$$

However, this is for a fixed B, while the conclusion of the lemma is that a.s. (A.4) holds simultaneously for every λ-null set $B \subset \mathfrak{X}$. There is in general an uncountable number of λ-null sets $B \subset \mathfrak{X}$, and we do not see how to use the argument in (A.4) to prove the result.

Instead, we argue as follows. First, by if necessary changing $\mu(y,\cdot)$ on a $\mathcal{L}(Y)$-null set of y, we may assume that

$$\mu(y,\cdot) \ll \lambda \qquad \text{for every } y \in \mathfrak{Y}. \tag{A.5}$$

Let $\nu(z,\cdot)$ be the regular conditional distribution of Y given Z. Then, for any $B \in \mathcal{X}$, a.s., using (A.1) and (A.2),

$$\begin{aligned}\mathbb{P}[X \in B \mid Z] = \mathbb{E}\left[\mathbf{1}_B(X) \mid Z\right] = \mathbb{E}\big[\mathbb{E}\left[\mathbf{1}_B(X) \mid Y\right] \mid Z\big] = \mathbb{E}\left[\mu(Y,B) \mid Z\right]\\ = \int_{\mathfrak{Y}} \mu(y,B)\nu(Z,\mathrm{d}y).\end{aligned} \tag{A.6}$$

This shows that (a version of) the regular conditional distribution μ' is given by the composition of the kernels ν and μ defined by

$$\mu'(z,B) := \int_{\mathfrak{Y}} \nu(z,\mathrm{d}y)\mu(y,B), \qquad B \in \mathcal{X}; \tag{A.7}$$

note that this composition is a probability kernel, see e.g. the more general [32, Lemma 1.41(iii)].

Now, if $\lambda(B) = 0$, then (A.5) shows that $\mu(y,B) = 0$ for every y, and hence (A.7) yields $\mu'(z,B) = 0$ for every $z \in \mathfrak{Z}$. Consequently, $\mu'(z,\cdot) \ll \lambda$ for every $z \in \mathfrak{Z}$. □

Lemma A.3. *Let X and Y be random variables taking values in Borel spaces $(\mathfrak{X},\mathcal{X})$ and $(\mathfrak{Y},\mathcal{Y})$, respectively. Let λ and λ' be σ-finite measures on $\mathfrak{X}$ and $\mathfrak{Y}$, respectively. Suppose that $\mathcal{L}(Y) \ll \lambda'$ and that the regular conditional distribution $\mu(y,\cdot)$ of X given Y satisfies $\mu(y,\cdot) \ll \lambda$ for $\mathcal{L}(Y)$-a.e. $y \in \mathfrak{Y}$. Then the distribution of (X,Y) in $\mathfrak{X}\times\mathfrak{Y}$ is absolutely continuous with respect to $\lambda\times\lambda'$.*

Proof. By if necessary changing $\mu(y,\cdot)$ on a $\mathcal{L}(Y)$-null set of $y \in \mathfrak{Y}$, we may assume that (A.5) holds.

Let $B \subset \mathfrak{X}\times\mathfrak{Y}$ with $\lambda\times\lambda'(B) = 0$. For $y \in \mathfrak{Y}$, let $B_y := \{x \in \mathfrak{X} : (x,y) \in B\}$. Let $A := \{y \in \mathfrak{Y} : \lambda(B_y) > 0\}$. By Fubini's theorem,

$$0 = \lambda\times\lambda'(B) = \int_{\mathfrak{X}\times\mathfrak{Y}} \mathbf{1}_B(x,y)\,\mathrm{d}\lambda(x)\,\mathrm{d}\lambda'(y) = \int_{\mathfrak{Y}} \lambda(B_y)\,\mathrm{d}\lambda'(y) \tag{A.8}$$

and thus $\lambda(B_y) = 0$ for λ'-a.e. y, i.e., $\lambda'(A) = 0$. Since $\mathcal{L}(Y) \ll \lambda'$, this implies

$$\mathbb{P}(Y \in A) = 0. \tag{A.9}$$

Furthermore, by (A.3),

$$\begin{aligned}\mathbb{P}[(X,Y) \in B] = \mathbb{E}\,\mathbf{1}_B(X,Y) = \mathbb{E}\int_{\mathfrak{X}} \mathbf{1}_B(x,Y)\mu(Y,\mathrm{d}x) = \mathbb{E}\int_{\mathfrak{X}} \mathbf{1}_{B_Y}(x)\mu(Y,\mathrm{d}x)\\ = \mathbb{E}\,\mu(Y,B_Y).\end{aligned} \tag{A.10}$$

If $Y \notin A$, then $\lambda(B_Y) = 0$, and thus $\mu(y,B_Y) = 0$ for every y by (A.5); in particular $\mu(Y,B_Y) = 0$. By (A.9), this shows that $\mu(Y,B_Y) = 0$ a.s., and thus (A.10) yields $\mathbb{P}[(X,Y) \in B] = 0$. □

For easy reference, we state also an elementary result on absolute continuity in $\mathbb{R}^d$, as in the main part of the paper this tacitly means with respect to Lebesgue measure.

Lemma A.4. *Let $T : \mathbb{R}^n \to \mathbb{R}^m$ be a linear operator, where $1 \leqslant m \leqslant n$. If $\mathbf{X}$ is a random vector in $\mathbb{R}^n$ with an absolutely continuous distribution, and T is onto, then the distribution of $T(\mathbf{X})$ in $\mathbb{R}^m$ is absolutely continuous.*

Proof. By changes of bases, we may assume that T is the projection to the first m coordinates. Let λ_d denote the Lebesgue measure in $\mathbb{R}^d$. If $A \subset \mathbb{R}^m$ with $\lambda_m(A) = 0$, then

$$\mathbb{P}\bigl(T(\mathbf{X}) \in A\bigr) = \mathbb{P}\bigl(\mathbf{X} \in A \times \mathbb{R}^{n-m}\bigr) = 0, \tag{A.11}$$

since $\lambda_n(A \times \mathbb{R}^{n-m}) = 0$. □

Appendix B. L^p theory

The proofs in this paper frequently use martingales and L^2 theory, in particular the identity (2.17). In this appendix, we extend the results to L^p estimates for any $p > 1$ by combining the arguments in Section 3 with the Burkholder–Davis–Gundy inequalities (see e.g. [32, Theorem 26.12]), which say that if $p \geqslant 1$, then there exist constants $c = c(p)$ and $C = C(p)$ such that for every (continuous-time) local martingale $M(t)$

$$c\,\mathbb{E}\,[M,M]_t^{p/2} \leqslant \mathbb{E}\,M^*(t)^p \leqslant C\,\mathbb{E}\,[M,M]_t^{p/2}, \qquad 0 \leqslant t \leqslant \infty. \tag{B.1}$$

(All constants in this appendix may depend on the exponent p.) We will mainly use the second inequality.

This extension to L^p leads to two major results. Using the case $p > 2$, we will obtain a proof of Theorem 12.3 (and therefore Theorem 12.5) showing convergence in L^p and thus moment convergence under natural conditions. Moreover, using the case $1 < p < 2$, we show that, as said in Remark 2.1, our main results hold also if we weaken the L^2 condition (A4) to L^p for some $p > 1$. More precisely, we will show the following.

Theorem B.1. *Theorems 1.8 and 4.1, and their extensions Theorems 8.4–8.6, all hold also if* (A4) *is replaced by the weaker*

(A4p) $\mathbb{E}\,|\xi_{ij}|^p < \infty$ *for all* $i, j \in \mathsf{Q}$ *and some* $p > 1$.

Also other results in this paper, for example the results on the drawn colours in Section 11, hold if (A4) is replaced by (A4p), provided we replace any L^2-norms by L^p norms $\|\ \|_p$; see also Remark 12.4 for the results on moments in Section 12. We leave the details to the reader.

We will basically follow the arguments in the main part of the paper, replacing L^2 estimates by L^p estimates, but sometimes the details of the arguments will differ. Moreover, we have chosen to first focus on obtaining the L^p estimates, leading to the proof of Theorem 12.3 (partly because this seems to be of greater interest for applications); we then return to the arguments yielding a.s. convergence and Theorem B.1. As before, we argue in several steps.

B.1. A single colour not influenced by others.

We begin with a colour i that is not influenced by any other (i.e., $i \in \mathsf{Q}_{\min}$), and prove an L^p-version of Lemmas 3.3 and 8.9.

Lemma B.2. *Assume* (A1)–(A3), (A5′) *(or* (A5)*), and* (A4p) *for some* $p > 1$*. Let* $i \in \mathsf{Q}_{\min}$*, and assume*

$$\textit{either} \quad i \notin \mathsf{Q}^- \ \textit{(i.e., } \xi_{ii} \geqslant 0 \textit{ a.s.)} \quad \textit{or} \quad \lambda_i > 0. \tag{B.2}$$

Then

(i) *The martingale* $e^{-\lambda_i t}X_i(t)$ *is* L^p*-bounded, and thus the a.s. limit* (3.13) *holds for some limit* $\mathcal{X}_i$*, and*

$$\widetilde{X}_i^{**} := \sup_{t\geqslant 0}\bigl\{e^{-\lambda_i t}X_i(t)\bigr\} \in L^p. \tag{B.3}$$

(ii) *Let* $(T_k)_1^\infty$ *be the times a ball of colour* i *is drawn, and let* η_k *be the number of balls of colour* i *that are added at time* T_k*. Let* $0 < q \leqslant p$*, and let* $f : \mathbb{R} \to \mathbb{R}$ *be a function such that* $\mathbb{E}\,|f(\xi_{ii})|^q < \infty$*. Finally, let* $\mu > 0$ *be such that* $(1 \wedge q)\mu > \lambda_i$*. Then (with a.s. convergent sum)*

$$\sum_{k=1}^\infty e^{-\mu T_k} f(\eta_k) \in L^q. \tag{B.4}$$

The statement in (i) that the martingale is L^p-bounded, is (since we have $p > 1$) by Doob's inequality equivalent to (B.3), but we state both for emphasis. Moreover, the statements are equivalent to $\mathcal{X}_i := \lim_{t\to\infty} e^{-\lambda_i t}X_i(t) \in L^p$. Note also that the definition of $\widetilde{X}_i^{**}$ in (B.3) agrees with (3.30) since $\kappa_i = 0$ and $\lambda_i^* = \lambda_i$ when $i \in \mathsf{Q}_{\min}$. Similarly, (B.2) is equivalent to (A7) for i, but for later use we prefer the form (B.2).

Proof. If $\lambda_i \leqslant 0$, then by (B.2) we have $i \notin \mathsf{Q}^-$ and thus $\xi_{ii} \geqslant 0$ a.s.; consequently $\xi_{ii} = 0$ a.s. and $X_i(t)$ is constant so the results are trivial (with an empty sum in (B.4)). We thus assume $\lambda_i > 0$.

Step 1: (ii) *for* $q = 1$. Since $\eta_k \overset{\mathrm{d}}{=} \xi_{ii}$ and is independent of T_k, we have by Lemma 3.2(ii), letting $t \to \infty$,

$$\mathbb{E}\sum_{k=1}^\infty e^{-\mu T_k}\bigl|f(\eta_k)\bigr| = a_i\,\mathbb{E}\,|f(\xi_{ii})| \int_0^\infty e^{-\mu s}\,\mathbb{E}\,X_i(s)\,\mathrm{d}s, \tag{B.5}$$

which is finite by (3.11) and the assumption $\mu > \lambda_i$.

Step 2: (ii) *for* $q < 1$. We have

$$\Bigl(\sum_{k=1}^\infty e^{-\mu T_k}\bigl|f(\eta_k)\bigr|\Bigr)^q \leqslant \sum_{k=1}^\infty e^{-q\mu T_k}\bigl|f(\eta_k)\bigr|^q \in L^1, \tag{B.6}$$

by Step 1 applied to $|f|^q$ and $q\mu$.

Step 3: If (i) *holds for some* $p > 1$*, then* (ii) *holds for all* $q \leqslant p$. We have already proved the case $q \leqslant 1$, so we may assume $q > 1$. In particular, $\mathbb{E}\,|f(\xi_{ii})| < \infty$. Furthermore, by induction (on $\lceil \log_2 q\rceil$), we may assume that (ii) holds if q is replaced by $q/2$.

We consider first two special cases, and then the general one.

(i) $\mathbb{E}\,f(\xi_{ii}) = 0$. In this case,

$$M(t) := \sum_{k=1}^\infty \mathbf{1}_{\{T_k\leqslant t\}} e^{-\mu T_k} f(\eta_k) \tag{B.7}$$

is a local martingale with $M(0) = 0$, since each T_k is a stopping time and $f(\eta_k)$ has mean 0 and is independent of $\mathcal{F}_{T_k}$. (Cf. Z_2 in the proof of Lemma 3.5.) The quadratic variation is by (2.16) (cf. (3.59))

$$[M, M]_t = \sum_{k=1}^\infty \mathbf{1}_{\{T_k\leqslant t\}} e^{-2\mu T_k} f(\eta_k)^2. \tag{B.8}$$

We have $f(\xi_{ii})^2 \in L^{q/2}$, so by the induction hypothesis, we have $[M, M]_\infty \in L^{q/2}$. (If $q/2 < 1$, note that $(q/2)2\mu = q\mu > \mu > \lambda_i$.) Consequently, (B.1) shows that M is an L^q-bounded martingale, which yields (B.4).

(ii) *$f = c$ is a constant.* It suffices to consider the case $c = 1$. We now define

$$M(t) := \sum_{k=1}^{\infty} \mathbf{1}_{\{T_k \leqslant t\}} e^{-\mu T_k} - \int_0^t e^{-\lambda_i s} a_i X_i(s) \, ds \tag{B.9}$$

and note that $M(t)$ is a local martingale (Cf. Z_3 in the proof of Lemma 3.5 and (3.55).) The quadratic variation is (cf. (3.57))

$$[M, M]_t = \sum_{k=1}^{\infty} \mathbf{1}_{\{T_k \leqslant t\}} e^{-2\mu T_k}. \tag{B.10}$$

Hence, in this case too, the induction hypothesis yields $[M, M]_\infty \in L^{q/2}$, and thus, (B.1) shows that $M(t)$ is an L^q-bounded martingale. Furthermore,

$$\sum_{k=1}^{\infty} e^{-\mu T_k} = M(\infty) + \int_0^\infty e^{-\mu s} a_i X_i(s) \, ds. \tag{B.11}$$

By assumption, (i) holds, and thus $\widetilde{X}_i^{**} \in L^p \subseteq L^q$ by (B.3). Furthermore, since $\mu > \lambda_i$,

$$\int_0^\infty e^{-\mu s} X_i(s) \, ds \leqslant \int_0^\infty e^{-\mu s + \lambda_i s} \widetilde{X}_i^{**} \, ds = (\mu - \lambda_i)^{-1} \widetilde{X}_i^{**} \in L^q. \tag{B.12}$$

This and (B.11) show that $\sum_k e^{-\mu T_k} \in L^q$, which is (B.4) in this case.

(iii) *General f.* We use the decomposition $f(x) = \big(f(x) - \mathbb{E} f(\xi_{ii})\big) + \mathbb{E} f(\xi_{ii})$ and the two preceding cases.

Step 4: (i) holds for all $p > 1$. By induction (on $\lceil \log_2 p \rceil$), we may for $p > 2$ assume that (i) holds if p is replaced by $p/2$.

As in the proof of Lemma 3.3, we let $M(t) := e^{-\lambda_i t} X_i(t)$, so that $M(t)$ is a martingale with quadratic variation (3.18):

$$[M, M]_t = X_i(0)^2 + \sum_{k=1}^{\infty} \mathbf{1}_{\{T_k \leqslant t\}} e^{-2\lambda_i T_k} \eta_k^2. \tag{B.13}$$

If $1 < p \leqslant 2$, we apply (ii) with $q = p/2 \leqslant 1$; this case holds by Step 1 or Step 2. If $p > 2$, we apply (ii) with $q = p/2$ and p replaced by $p/2 > 1$; this case holds by the induction hypothesis and Step 3. In both cases, we take $f(x) = x^2$ and note that $\mathbb{E} |f(\xi_{ii})|^q = \mathbb{E} |\xi_{ii}|^p < \infty$. Hence, taking $\mu := 2\lambda_i$, (B.4) shows that $[M, M]_\infty \in L^{p/2}$. Consequently, (B.1) shows that M is an L^p-bounded martingale and that (B.3) holds. This completes the proof. ☐

We will also need a quantitative version of Lemma B.2(i).

Lemma B.3. *Under the assumptions of Lemma B.2, we have*

$$\Big\| \sup_{t \geqslant 0} \big| e^{-\lambda_i t} X_i(t) \big| \Big\|_p \leqslant C \lceil X_i(0) \rceil, \tag{B.14}$$

where C does not depend on $X_i(0)$.

Proof. It ought to be straightforward to keep track of the norms of all quantities in the proof of Lemma B.2, but it seems simpler to argue as follows. First, if $X_i(0) = m$ is an integer, then the process $X_i(t)$ can be seen as the sum of m independent and identically distributed processes $X_i^{(k)}(t)$, $k = 1, \dots, m$, each started with $X_i^{(k)}(0) = 1$ (but otherwise the same as $X_i(t)$). Hence,

$$\sup_{t\geqslant 0}\bigl|e^{-\lambda_i t}X_i(t)\bigr| \leqslant \sum_{k=1}^{m}\sup_{t\geqslant 0}\bigl|e^{-\lambda_i t}X_i^{(k)}(t)\bigr|, \tag{B.15}$$

and (B.14) follows by (B.3) and Minkowski's inequality.

If $X_i(0)$ is not an integer, let $X_i'(t)$ be an independent copy of $X_i(t)$ started with $X_i'(0) = \lceil X_i(0)\rceil - X_i(0)$. Then $X_i(t) + X_i'(t)$ is a copy of the same process started with $\lceil X_i(0)\rceil$. Since $0 \leqslant X_i(t) \leqslant X_i(t) + X_i'(t)$, the result follows from the special case just treated. ☐

B.2. **A colour only produced by one other colour.** Consider now the situation in Lemmas 3.5 and 8.11, with two colours i and j such that $\mathsf{P}_i = \{j\}$, and also $X_i(0) = 0$. We fix such i and j throughout this subsection. (We may repeat the assumptions for emphasis.)

As in Section 3.2, let $0 < T_1 < T_2 < \dots$ be the times that a ball of colour j is drawn, and let $\mathcal{F}_{T_k}$ be the corresponding σ-fields.

Recall the notation (3.30). Define also, for $\mu \in \mathbb{R}$,

$$\kappa^*(\mu) := \begin{cases} \kappa_j, & \mu < \lambda_j^*, \\ \kappa_j + 1, & \mu = \lambda_j^*, \\ 0, & \mu > \lambda_j^*. \end{cases} \tag{B.16}$$

Note that, by (2.9), $\kappa_i = \kappa^*(\lambda_i)$. (We are mainly interested in the case $\mu = \lambda_i$, but we use induction to prove Lemma B.5 below, and we will then need more general μ.)

Lemma B.4. *Let $\mu \in \mathbb{R}$ and define*

$$V(t) := \int_0^t e^{-\mu s}X_j(s)\,\mathrm{d}s. \tag{B.17}$$

Then

$$\widetilde{V}^{**} := \sup_{t\geqslant 0}\bigl\{(t+1)^{-\kappa^*(\mu)}e^{-(\lambda_j^*-\mu)_+t}|V(t)|\bigr\} \leqslant C\widetilde{X}_j^{**}. \tag{B.18}$$

Proof. By (3.30) we have, considering the three cases in (B.16) separately,

$$V(t) \leqslant \int_0^t (s+1)^{\kappa_j}e^{(\lambda_j^*-\mu)s}\widetilde{X}_j^{**}\,\mathrm{d}s \leqslant C\widetilde{X}_j^{**}(t+1)^{\kappa^*(\mu)}e^{(\lambda_j^*-\mu)_+t}, \tag{B.19}$$

and (B.18) follows. ☐

Lemma B.5. *Assume* (A1)–(A3), (A5′) *(or* (A5)*), and* (A4p) *for some $p > 1$. Suppose that $i, j \in \mathsf{Q}$ are such that $\mathsf{P}_i = \{j\}$ and $X_i(0) = 0$, and suppose also that*

$$\widetilde{X}_j^{**} \in L^p. \tag{B.20}$$

Let $(\zeta_k)_1^\infty$ be a sequence of random variables with the same distribution such that ζ_k is independent of $\mathcal{F}_{T_k}$. Let $\mu \in \mathbb{R}$ be such that

$$\begin{cases} \mu \geqslant 0, & \text{if } i \notin \mathsf{Q}^-, \\ \mu \vee \lambda_j^* > 0, & \text{if } i \in \mathsf{Q}^-, \end{cases} \tag{B.21}$$

and let

$$Z(t) := \sum_{k=1}^{\infty} \mathbf{1}_{\{T_k \leqslant t\}} e^{-\mu T_k} \zeta_k. \tag{B.22}$$

If $1 < q \leqslant p$ *and* $\mathbb{E}\,|\zeta_k|^q < \infty$*, then*

$$\widetilde{Z}^{**} := \sup_{t \geqslant 0} \bigl\{(t+1)^{-\kappa^*(\mu)} e^{-(\lambda_j^* - \mu)_+ t} |Z(t)|\bigr\} \in L^q. \tag{B.23}$$

Proof. The proof is similar to Step 5 in the proof of Lemma 3.5 (which essentially is the case $q = 2$ of the present lemma), now using (B.1).

By induction (on $\lceil \log_2 q \rceil$), we may for $q > 2$ assume that the lemma holds for $q/2$.

Case 1: $\mathbb{E}\,\zeta_k = 0$. In this case, $Z(t)$ is a local martingale, for the same reason as $M(t)$ in (B.7). Its quadratic variation is, by (2.16) again,

$$[Z, Z]_t = \sum_{k=1}^{\infty} \mathbf{1}_{\{T_k \leqslant t\}} e^{-2\mu T_k} |\zeta_k|^2. \tag{B.24}$$

We consider two subcases.

(i) $1 < q \leqslant 2$. By (B.1), together with (B.24) and the independence of ζ_k and T_k, we have, since $q/2 \leqslant 1$,

$$\begin{aligned}
\mathbb{E}\, Z^*(t)^q &\leqslant C\,\mathbb{E}\,[Z, Z]_t^{q/2} = C\,\mathbb{E}\Bigl(\sum_{k=1}^{\infty} \mathbf{1}_{\{T_k \leqslant t\}} e^{-2\mu T_k} |\zeta_k|^2\Bigr)^{q/2} \\
&\leqslant C\,\mathbb{E} \sum_{k=1}^{\infty} \mathbf{1}_{\{T_k \leqslant t\}} e^{-q\mu T_k} |\zeta_k|^q.
\end{aligned} \tag{B.25}$$

Hence, Lemma 3.2(ii) and (3.30) yield

$$\mathbb{E}\, Z^*(t)^q \leqslant C a_j\, \mathbb{E} \int_0^t e^{-q\mu s} X_j(s)\,\mathrm{d}s \leqslant C\,\mathbb{E}\,\widetilde{X}_j^{**} \int_0^t (s+1)^{\kappa_j} e^{(\lambda_j^* - q\mu)s}\,\mathrm{d}s. \tag{B.26}$$

In the sequel, we allow constants C to depend on $\|\widetilde{X}_j^{**}\|_p$ (which is finite by assumption); hence we may absorb $\mathbb{E}\,\widetilde{X}_j^{**}$ into C in (B.26).

We consider three subsubcases:

(i)(a) $\lambda_j^* < q\mu$. In this case, we may take $t = \infty$ in (B.26) and obtain $Z^*(\infty) \in L^q$. The result (B.23) follows since $\widetilde{Z}^{**} \leqslant Z^*(\infty)$.

(i)(b) $\lambda_j^* \geqslant q\mu$ *and* $\lambda_j^* > 0$. Define, similarly to (3.70),

$$\widetilde{Z}^{\dagger}(n) := \sup_{n-1 \leqslant t \leqslant n} (t+1)^{-\kappa^*(\mu)} e^{-(\lambda_j^* - \mu)t} |Z(t)| \leqslant C n^{-\kappa^*(\mu)} e^{-(\lambda_j^* - \mu)n} Z^*(n). \tag{B.27}$$

By (B.26), we thus have

$$\begin{aligned}
\mathbb{E}\,\widetilde{Z}^{\dagger}(n)^q &\leqslant C n^{-q\kappa^*(\mu)} e^{-q(\lambda_j^* - \mu)n} \int_0^n (s+1)^{\kappa_j} e^{(\lambda_j^* - q\mu)s}\,\mathrm{d}s \\
&\leqslant C n^{1+\kappa_j - q\kappa^*(\mu)} e^{(-q(\lambda_j^* - \mu) + \lambda_j^* - q\mu)n} = C n^{1+\kappa_j - q\kappa^*(\mu)} e^{-(q-1)\lambda_j^* n}.
\end{aligned} \tag{B.28}$$

We have $(q-1)\lambda_j^* > 0$, and thus (B.28) implies

$$\mathbb{E}\,(\widetilde{Z}^{**})^q \leqslant \mathbb{E} \sum_{n=1}^{\infty} \widetilde{Z}^{\dagger}(n)^q < \infty, \tag{B.29}$$

which shows (B.23).

(i)(c) $\lambda_j^* \geqslant q\mu$ *and* $\lambda_j^* \leqslant 0$. Then also $\mu \leqslant 0$, and thus $\mu \vee \lambda_j^* \leqslant 0$. By (B.21), we must have $i \notin \mathsf{Q}^-$ and $\mu = 0$, and then also $\lambda_j^* = 0$. Hence, $\kappa(\mu) = \kappa_j + 1$ by (B.16). We define, similarly to (3.74),

$$\widetilde{Z}^{\ddagger}(n) := \sup_{2^{n-1} \leqslant t \leqslant 2^n} t^{-\kappa^*(\mu)} e^{-(\lambda_j^* - \mu)t} |Z(t)| = \sup_{2^{n-1} \leqslant t \leqslant 2^n} t^{-\kappa_j - 1} |Z(t)|. \tag{B.30}$$

Similarly to (B.28), it follows from (B.26) that

$$\mathbb{E}\, \widetilde{Z}^{\ddagger}(n)^q \leqslant C 2^{-q(\kappa_j+1)n} \int_0^{2^n} (s+1)^{\kappa_j}\, \mathrm{d}s \leqslant C 2^{(1-q)(\kappa_j+1)n} \leqslant C 2^{-(q-1)n}. \tag{B.31}$$

Hence, (B.23) follows by

$$\mathbb{E}\, (\widetilde{Z}^{**})^q \leqslant \mathbb{E}\, Z^*(1)^q + \mathbb{E} \sum_{n=1}^{\infty} \widetilde{Z}^{\ddagger}(n)^q < \infty, \tag{B.32}$$

since $\mathbb{E}\, Z^*(1)^q < \infty$ by (B.26).

(ii) $q > 2$. By (B.1), together with (B.24) and the induction hypothesis (for $q/2$ and 2μ, using $\mathbb{E}\, |\zeta_k^2|^{q/2} < \infty$)

$$\mathbb{E}\, Z^*(t)^q \leqslant C\, \mathbb{E}\, [Z, Z]_t^{q/2} \leqslant C\bigl((t+1)^{\kappa^*(2\mu)} e^{(\lambda_j^* - 2\mu)_+ t}\bigr)^{q/2}. \tag{B.33}$$

Again, we consider three subsubcases:

(ii)(a) $\lambda_j^* < 2\mu$. Then $\kappa^*(2\mu) = 0$ by (B.16), and thus we may let $t \to \infty$ in (B.33) and obtain $\mathbb{E}\, Z^*(\infty)^q < \infty$, which yields (B.23) since $\widetilde{Z}^{**} \leqslant Z^*(\infty)$.

(ii)(b) $\lambda_j^* \geqslant 2\mu$ *and* $\lambda_j^* > 0$. Define again $\widetilde{Z}^{\dagger}(n)$ by (B.27). Then, by (B.33),

$$\begin{aligned} \mathbb{E}\, \widetilde{Z}^{\dagger}(n)^q &\leqslant C n^{-q\kappa^*(\mu) + \frac{q}{2}\kappa^*(2\mu)} e^{(-q(\lambda_j^* - \mu) + \frac{q}{2}(\lambda_j^* - 2\mu))n} \\ &\leqslant C n^{-q\kappa^*(\mu) + \frac{q}{2}\kappa^*(2\mu)} e^{-\frac{q}{2}\lambda_j^* n}. \end{aligned} \tag{B.34}$$

Hence, we have again (B.29), and thus (B.23).

(ii)(c) $\lambda_j^* \geqslant 2\mu$ *and* $\lambda_j^* \leqslant 0$. Again, by (B.21), we must have $i \notin \mathsf{Q}^-$ and $\mu = 0$, and then also $\lambda_j^* = 0$. Hence, $\kappa^*(\mu) = \kappa^*(2\mu) = \kappa_j + 1$. Define again $\widetilde{Z}^{\ddagger}(n)$ by (B.30). Then, by (B.33),

$$\mathbb{E}\, \widetilde{Z}^{\ddagger}(n)^q \leqslant C 2^{(-q(\kappa_j+1) + \frac{q}{2}\kappa^*(2\mu))n} = C 2^{-\frac{q}{2}(\kappa_j+1)n}. \tag{B.35}$$

Consequently, (B.32) holds, and thus (B.23).

Case 2: $\zeta_k = c$ is a constant. We may assume $c = 1$.

Let $V(t)$ be as in (B.17) and define

$$M(t) := Z(t) - a_j V(t) = \sum_{k=1}^{\infty} \mathbf{1}_{\{T_k \leqslant t\}} e^{-\mu T_k} - a_j \int_0^t e^{-\mu s} X_j(s)\, \mathrm{d}s. \tag{B.36}$$

Then $M(t)$ is a local martingale by the same argument as for (B.9) (i.e., as for Z_3 in the proof of Lemma 3.5), and its quadratic variation is (cf. (3.57))

$$[M, M]_t = \sum_{k=1}^{\infty} \mathbf{1}_{\{T_k \leqslant t\}} e^{-2\mu T_k}. \tag{B.37}$$

This is the same as in (B.24), except for the factor $|\zeta_k|^2$ there (or the same if we choose $\zeta_k = \pm 1$). Consequently, the argument in Case 1 yields

$$\widetilde{M}^{**} := \sup_{t\geqslant 0}\bigl\{(t+1)^{-\kappa^*(\mu)}e^{-(\lambda_j^*-\mu)_+t}|M(t)|\bigr\} \in L^q. \tag{B.38}$$

Furthermore, (B.18) and the assumption (B.20) yield $\widetilde{V}^{**} \in L^p \subseteq L^q$. Consequently, (B.23) follows by (B.36).

Case 3: General ζ_k. The result follows from Cases 1 and 2 by the decomposition $\zeta_k = (\zeta_k - \mathbb{E}\,\zeta_k) + \mathbb{E}\,\zeta_k$. □

Lemma B.6. *Assume* (A1)–(A3), (A5′) *(or* (A5)*), and* (A4p) *for some* $p > 1$. *Suppose that* $i, j \in \mathsf{Q}$ *are such that* $\mathsf{P}_i = \{j\}$ *and* $X_i(0) = 0$, *and suppose also that* (B.2) *holds. If* $\widetilde{X}_j^{**} \in L^p$, *then* $\widetilde{X}_i^{**} \in L^p$.

Proof. We use again the decomposition (3.28), where we recall that $Y_k(t)$ denote copies of the one-colour process in Section 3.1, and that the process $Y_k(t)$ is independent of $\mathcal{F}_{T_k}$. Let

$$\zeta_k := \sup_{t\geqslant 0}\bigl\{e^{-\lambda_i t}Y_k(t)\bigr\}. \tag{B.39}$$

Then (3.28) implies

$$\begin{aligned} e^{-\lambda_i t}X_i(t) &= \sum_{k=1}^{\infty}\mathbf{1}_{\{T_k\leqslant t\}}e^{-\lambda_i T_k}\cdot e^{-\lambda_i(t-T_k)}Y_k(t-T_k) \\ &\leqslant \sum_{k=1}^{\infty}\mathbf{1}_{\{T_k\leqslant t\}}e^{-\lambda_i T_k}\zeta_k =: Z(t). \end{aligned} \tag{B.40}$$

Let $\eta_k = Y_k(0)$, and recall that $\eta_k \overset{\mathrm{d}}{=} \xi_{ji} \in L^p$. Then, conditioning on η_k, Lemma B.3 applies to $Y_k(t)$ and shows that

$$\mathbb{E}\left(\zeta_k^p \mid \eta_k\right) \leqslant C\lceil\eta_k\rceil^p \leqslant C(\eta_k+1)^p. \tag{B.41}$$

Consequently, using (A4p),

$$\mathbb{E}\,\zeta_k^p \leqslant C\,\mathbb{E}\,(\eta_k+1)^p = \mathbb{E}\,(\xi_{ji}+1)^p < \infty. \tag{B.42}$$

We apply Lemma B.5 to the sum $Z(t)$ in (B.40), taking $q = p$ and $\mu = \lambda_i$. Note that then (B.21) holds: if $i \notin \mathsf{Q}^-$ then $\xi_{ii} \geqslant 0$ a.s., and thus $\mu = \lambda_i \geqslant 0$; if $i \in \mathsf{Q}^-$ then $\mu = \lambda_i > 0$ by our assumption (B.2). Hence, (B.23) holds. Furthermore, $\kappa^*(\mu) = \kappa_i$ and $(\lambda_j^* - \mu)_+ = (\lambda_i^* - \lambda_i)$, as is easily verified by considering the three cases in (B.16) (and after (3.40)) separately. Hence, (3.30), (B.40), and (B.23) yield

$$\widetilde{X}_i^{**} = \sup_{t\geqslant 0}\bigl\{(t+1)^{-\kappa_i}e^{-(\lambda_i^*-\lambda_i)t}\cdot e^{-\lambda_i t}X_i(t)\bigr\} \leqslant \widetilde{Z}^{**} \in L^p. \tag{B.43}$$

□

B.3. **The general case for a single colour.** We now consider any colour $i \in \mathsf{Q}$.

Lemma B.7. *Assume* (A1)–(A3), *either* $\bigl($(A5′) *and* (A7)$\bigr)$ *or* (A5), *and* (A4p) *for some* $p > 1$. *Let* $i \in \mathsf{Q}$, *and assume that for every* $j \in \mathsf{P}_i$, *we have* $\widetilde{X}_j^{**} \in L^p$. *Then* $\widetilde{X}_i^{**} \in L^p$.

Proof. Since (A5) implies (A5′) and (A7), these hold in any case. We consider two cases.

Case 1: (B.2) *holds, i.e.* $i \notin \mathsf{Q}^-$ *or* $\lambda_i > 0$. As in the proof of Lemmas 3.7 and 8.14, we split the colour i into subcolours i_0 and i_j, $j \in \mathsf{P}_i$. We then use Lemma B.2(i) for i_0 and Lemma B.6 for every i_j, and the result follows by (3.91).

Case 2: $i \in \mathsf{Q}^-$ *and* $\lambda_i \leqslant 0$. Then (A7) yields $\lambda_i^* > 0$. Hence $\lambda_i^* > \lambda_i$, and since $\lambda_i^* = \lambda_i \vee \max_{j \in \mathsf{P}_i} \lambda_j^*$, we have $\mathsf{P}_i \neq \varnothing$ and

$$\lambda_i^* = \max_{j \in \mathsf{P}_i} \lambda_j^*. \tag{B.44}$$

Consider a modification $\bar{\xi}_{ii}$ of ξ_{ii} such that $\bar{\xi}_{ii} \geqslant \xi_{ii}$ a.s., and $\bar{\lambda}_i := a_i \mathbb{E}\,\bar{\xi}_{ii} \in (0, \lambda_i^*)$. (For example, let $\bar{\xi}_{ii} := \xi_{ii} \vee \tilde{\xi}$ where $\tilde{\xi} \in \{\pm 1\}$ is independent of ξ_{ii} and $\mathbb{P}(\tilde{\xi} = 1) \in (0,1]$ is chosen suitably.) Modify the urn by replacing ξ_{ii} by $\bar{\xi}_{ii}$; this does not affect any colour $j \prec i$; in particular $X_j(t)$ and all draws of colour j remain the same for every $j \in \mathsf{P}_i$, but at each draw of colour i we may add more balls of colour i; hence, letting $\overline{X}_i(t)$ denote the number of balls of colour i in the modified urn, we have, using an obvious coupling of the two urns,

$$\overline{X}_i(t) \geqslant X_i(t), \qquad t \geqslant 0. \tag{B.45}$$

The modified urn satisfies all our conditions in the present lemma, and since $\bar{\lambda}_i > 0$, the already proven Case 1 shows that (with obvious notation) $\widetilde{\overline{X}}_i^{**} \in L^p$. Furthermore, $\bar{\lambda}_i < \lambda_i^*$, and thus (B.44) implies

$$\bar{\lambda}_i^* := \bar{\lambda}_i \vee \max_{j \in \mathsf{P}_i} \lambda_j^* = \bar{\lambda}_i \vee \lambda_i^* = \lambda_i^*. \tag{B.46}$$

Similarly, using (2.9), we have $\bar{\kappa}_i = \kappa_i$. Consequently, the exponents in (3.30) are the same for $X_i(t)$ and $\overline{X}_i(t)$, and thus (B.45) implies

$$\widetilde{X}_i^{**} \leqslant \widetilde{\overline{X}}_i^{**}. \tag{B.47}$$

Since, as just shown, $\widetilde{\overline{X}}_i^{**} \in L^p$, this completes the proof. □

B.4. L^p bounds and convergence.

Lemma B.8. *Assume* (A1)–(A3)*, either* ((A5′) *and* (A7)) *or* (A5)*, and* (A4p) *for some* $p > 1$*. Then* $\widetilde{X}_i^{**} \in L^p$ *for every* $i \in \mathsf{Q}$*.*

Proof. By Lemma B.7 and induction on the colour i. □

Proof of Theorem 12.3. By Lemma B.8, $\widetilde{X}_j^{**} \in L^p$; hence it follows, exactly as for the case $p = 2$ in Theorem 12.2, that the collection $\{|\widetilde{X}_i(t)|^p : t \geqslant 1\}$ is uniformly integrable, and thus the convergence (4.1) holds also in L^p. □

Once we have proved Theorem B.1, the proof of Theorem 12.3 applies to any $p > 1$, as claimed in Remark 12.4.

B.5. **A.s. convergence.** We now turn Theorem B.1, i.e., that our a.s. convergence results hold also if (A4) is replaced by (A4p). We may assume $1 < p < 2$, since for $p \geqslant 2$ the assumption (A4p) implies (A4), and the results are already proven.

We begin by extending Lemma 8.9 to $1 < p < 2$, complementing Lemmas B.2 and B.3 for the case excluded there when (B.2) does not hold.

Lemma B.9. *Assume* (A1)–(A3), (A5′), *and* (A4p) *for some* $p \in (1, 2]$. *Let* $i \in \mathsf{Q}_{\min}$ *and assume* $\lambda_i \leqslant 0$. *Let* $x_0 := X_i(0)$.

(i) *If* $\lambda_i = 0$, *then* $X_i(t)$ *is a martingale with*

$$X_i(t) \xrightarrow{\text{a.s.}} \mathcal{X}_i = 0, \qquad \text{as } t \to \infty, \tag{B.48}$$

$$\mathbb{E}\, X_i^*(t)^p \leqslant C x_0^p + C x_0 t, \qquad \text{for every } t < \infty. \tag{B.49}$$

Furthermore, for every $\delta > 0$,

$$\mathbb{E}\,\bigl(\sup_{t\geqslant 0}\{e^{-\delta t} X_i(t)\}\bigr)^p < \infty. \tag{B.50}$$

(ii) *If* $\lambda_i < 0$, *then, with* $M(t) := e^{-\lambda_i t} X_i(t)$,

$$X_i(t) \xrightarrow{\text{a.s.}} \mathcal{X}_i = 0, \qquad \text{as } t \to \infty, \tag{B.51}$$

$$\mathbb{E}\, X_i(t)^p \leqslant C x_0^p e^{\lambda_i t}, \tag{B.52}$$

$$\mathbb{E}\, M^*(t)^p \leqslant C x_0^p e^{-(p-1)\lambda_i t}. \tag{B.53}$$

Proof. In both cases, $X_i(t)$ is a (sub)critical continuous-time branching process and therefore a.s. dies out, see Remark 8.2, which gives (B.48) and (B.51).

Recall from (3.12) that $M(t) := e^{-\lambda_i t} X_i(t)$ is a martingale (this does not require (A4), only $\mathbb{E}\,\xi_{ii} < \infty$), and hence, or by (3.11),

$$\mathbb{E}\, X_i(t) = e^{\lambda_i t}\, \mathbb{E}\, M(t) = e^{\lambda_i t} M(0) = x_0 e^{\lambda_i t}. \tag{B.54}$$

It follows from (3.18) that

$$[M, M]_t^{p/2} \leqslant x_0^p + \sum_{k=1}^{\infty} \mathbf{1}_{\{T_k \leqslant t\}} e^{-p\lambda_i T_k} \eta_k^p, \tag{B.55}$$

and hence, using Lemma 3.2(ii) and (B.54),

$$\mathbb{E}\,[M, M]_t^{p/2} \leqslant x_0^p + C \int_0^t e^{-p\lambda_i s}\, \mathbb{E}\, X_i(s)\,\mathrm{d}s = x_0^p + C x_0 \int_0^t e^{-(p-1)\lambda_i s}\,\mathrm{d}s. \tag{B.56}$$

If $\lambda_i = 0$, this yields (B.49) by (B.1); then (B.50) follows as in (8.12).

If $\lambda_i < 0$, then (B.56) and (B.1) yield, recalling that x_0 is an integer,

$$\mathbb{E}\, M^*(t)^p \leqslant C x_0^p + C x_0 e^{-(p-1)\lambda_i t} \leqslant C x_0^p e^{-(p-1)\lambda_i t} \tag{B.57}$$

and thus

$$\mathbb{E}\, X_i(t)^p = e^{p\lambda_i t}\, \mathbb{E}\, M^*(t)^p \leqslant C x_0^p e^{\lambda_i t}, \tag{B.58}$$

showing (B.53) and (B.52). ☐

Proof of Theorem B.1. As said above, we may assume $1 < p < 2$. It suffices to prove the continuous-time versions Theorem 4.1 and 8.6; then the proofs of Theorems 1.8, 8.4, and 8.5 are as before.

By Lemma B.8, we have $\widetilde{X}_i^{**} \in L^p$ for every $i \in \mathsf{Q}$, but it remains to show that $\widetilde{X}_i(t)$ converges. We follow the proof of Theorem 4.1 (and Theorem 8.6) step by step in the claims below which extend the limit statements in the lemmas in Section 3 and Section 8 (recall that L^p versions of the L^2 estimates there already are given); we omit some details. Assumptions on $\|\widetilde{X}_j^{**}\|_2$ are replaced by $\|\widetilde{X}_j^{**}\|_p$ (and hold in our case by Lemma B.8). We assume in the sequel (A1)–(A3), (A5′) (or (A5)), and (A4p) for some $p \in (1, 2]$. (But not (A7) unless said so.)

Note first that (3.11)–(3.12) still hold. (In fact, they require only the first moment $\mathbb{E}\,\xi_{ii} < \infty$.)

(i) *In Lemmas 3.3 and 8.9, the convergence* $\widetilde{X}_i := e^{-\lambda_i t} X_i(t) \xrightarrow{\text{a.s.}} \mathcal{X}_i$ *still holds. We still have* $\mathcal{X}_i > 0$ *a.s. when* $i \notin \mathsf{Q}^-$, *and* $\mathbb{P}(\mathcal{X}_i > 0) > 0$ *when if* $i \in \mathsf{Q}^-$ *and* $\lambda_i > 0$.

Proof. The convergence $\widetilde{X}_i(t) \xrightarrow{\text{a.s.}} \mathcal{X}_i$ for some limit $\mathcal{X}_i \in L^p$ is shown in Lemma B.2 or B.9. Furthermore, if $i \notin \mathsf{Q}^-$, then the argument in the proof of Lemma 3.3 shows that $\mathcal{X}_i > 0$ a.s.; otherwise, if $\lambda_i > 0$, then $\mathbb{E}\,\mathcal{X}_i = X_i(0) > 0$ and thus at least $\mathbb{P}(\mathcal{X}_i > 0) > 0$. ☐

(ii) *In Lemmas 3.5 and 8.11, the convergence* $\widetilde{X}_i(t) \xrightarrow{\text{a.s.}} \mathcal{X}_i$ *still holds; furthermore,* $\mathcal{X}_j > 0 \implies \mathcal{X}_i > 0$ *a.s.*

Proof. We argue as in the proof of Lemmas 3.5 and 8.11, using again the decompositions (3.40) and (3.43). For $Z_4(t)$, the argument in (3.49)–(3.54) holds without changes. Also for $Z_3(t)$ the argument in the proof of Lemma 3.5 still holds, since the L^2 estimate (3.58) remains valid.

The remaining terms $Z_1(t)$ and $Z_2(t)$ are still local martingales. For $Z_2(t)$ we have by the definition (3.59), (B.1), Lemma 3.2(ii), and (3.30)

$$\begin{aligned}\mathbb{E}\,Z_2^*(t)^p &\leqslant C\,\mathbb{E}\,[Z_2, Z_2]_t^{p/2} \leqslant C\,\mathbb{E}\sum_{k=1}^{\infty} \mathbf{1}_{\{T_k \leqslant t\}} e^{-p\lambda_i T_k} |\eta_k - r_{ji}|^p \\ &= C\,\mathbb{E}\int_0^t e^{-p\lambda_i s} X_j(s)\,\mathrm{d}s \\ &\leqslant C\,\mathbb{E}\,\widetilde{X}_j^{**} \int_0^t (s+1)^{\kappa_j} e^{(\lambda_j^* - p\lambda_i)s}\,\mathrm{d}s. \end{aligned} \tag{B.59}$$

We now argue as in Step 5 of the proof of Lemma 3.5, using (B.59) instead of (3.68) and L^p instead of L^2; we replace 2 by p in all exponents, and the cases (i′) and (ii′) are replaced by $\lambda_j^* < p\lambda_i$ and ($\lambda_j^* \geqslant p\lambda_i$ and $\lambda_j^* > 0$). We obtain as before that $\widetilde{Z}_2(t) \xrightarrow{\text{a.s.}} \mathcal{Z}_2$ as $t \to \infty$, where the limit $\mathcal{Z}_2 = 0$ except in case (i).

For $Z_1(t)$ we use (3.64) and both directions of (B.1) to obtain

$$\mathbb{E}\,|Z_1^*(t)|^p \leqslant C\,\mathbb{E}\,[Z_1, Z_1]_t^{p/2} \leqslant C\,\mathbb{E}\sum_{k=1}^{\infty} [Z_1^{(k)}, Z_1^{(k)}]_t^{p/2} \leqslant C\,\mathbb{E}\sum_{k=1}^{\infty} |Z_1^{(k)*}(t)|^p. \tag{B.60}$$

The definition (3.62) yields

$$\bigl|Z_1^{(k)*}(t)\bigr|^p = \mathbf{1}_{\{t \geqslant T_k\}} e^{-p\lambda_i T_k} |\widetilde{Y}_k^*(t - T_k)|^p, \tag{B.61}$$

where the martingale $\widetilde{Y}_k(t) := e^{-\lambda_i t} Y_k(t) - Y_k(0)$ is independent of T_k, and $Y_k(t)$ is a copy of the one-colour process in Section 3.1 and Lemma B.9. Thus (B.60) and

Lemma 3.2(ii) yield

$$\mathbb{E}\,|Z_1^*(t)|^p \leqslant C\,\mathbb{E}\sum_{k=1}^{\infty} \mathbf{1}_{\{t\geqslant T_k\}} e^{-p\lambda_i T_k}\widetilde{Y}_k^*(t)^p$$
$$= C\,\mathbb{E}\big[\widetilde{Y}_1^*(t)^p\big]\,\mathbb{E}\int_0^t e^{-p\lambda_i s}X_j(s)\,\mathrm{d}s. \tag{B.62}$$

We separate three cases, as in the proof of Lemma 8.11:

(1) $i \notin \mathsf{Q}^-$ or $\lambda_i > 0$: This is condition (B.2); thus Lemma B.3 applies to $Y_k(t)$ conditionally on η_k, which gives
$$\mathbb{E}\big[\widetilde{Y}_k^*(t)^p \mid \eta_k\big] \leqslant C\lceil \eta_k\rceil^p, \tag{B.63}$$
and consequently
$$\mathbb{E}\big[\widetilde{Y}_1^*(t)^p\big] \leqslant C. \tag{B.64}$$
Thus (B.62) yields the same estimate (B.59) as for $Z_2^*(t)$, and the argument after (B.59) applies to Z_1 too and yields $\widetilde{Z}_1(t) \overset{\text{a.s.}}{\longrightarrow} \mathcal{Z}_1$ for some $\mathcal{Z}_1$, with $\mathcal{Z}_1 = 0$ except in case (i).

(2) $i \in \mathsf{Q}^-$ and $\lambda_i = 0$: (As in the proof of Lemma 8.11, the assumptions then yield $\lambda_j^* = \lambda_i^* > 0$.) In this case, $\widetilde{Y}_k(t) = Y_k(t) - Y_k(0)$ where $Y_k(t)$ is a copy of $X_i(t)$ in Lemma B.9. Thus, (B.49) yields, recalling that now η_k is an integer, for $t \geqslant 1$,
$$\mathbb{E}\big[\widetilde{Y}_k^*(t)^p \mid \eta_k\big] \leqslant C\eta_k^p t, \tag{B.65}$$
and consequently
$$\mathbb{E}\big[\widetilde{Y}_1^*(t)^p\big] \leqslant Ct. \tag{B.66}$$
This and (B.62) yield, still for $t \geqslant 1$,
$$\mathbb{E}\,|Z_1^*(t)|^p \leqslant Ct\,\mathbb{E}\int_0^t e^{-p\lambda_i s}X_j(s)\,\mathrm{d}s \leqslant Ct\,\mathbb{E}\,\widetilde{X}_j^{**}\int_0^t (s+1)^{\kappa_j}e^{(\lambda_j^*-p\lambda_i)s}\,\mathrm{d}s$$
$$\leqslant Ct^{\kappa_j+1}e^{(\lambda_j^*-p\lambda_i)t}. \tag{B.67}$$
This differs from (B.59) by an extra factor t, but the argument after (B.59) (in this case modifications of (3.70)–(3.73)) still works and yields $\widetilde{Z}_1(t) \overset{\text{a.s.}}{\longrightarrow} 0$.

(3) $\lambda_i < 0$: (The assumptions yield $\lambda_j^* = \lambda_i^* > 0$ in this case too.) In this case, we condition on T_k and η_k and then use (B.61) and (B.53), yielding
$$\mathbb{E}\big[|Z_1^{(k)*}(t)|^p \mid T_k,\eta_k\big] = \mathbf{1}_{\{t\geqslant T_k\}}e^{-p\lambda_i T_k}\,\mathbb{E}\big[|\widetilde{Y}_k^*(t-T_k)|^p \mid T_k,\eta_k\big]$$
$$\leqslant C\mathbf{1}_{\{t\geqslant T_k\}}e^{-p\lambda_i T_k}\eta_k^p e^{-(p-1)\lambda_i(t-T_k)}$$
$$= C\mathbf{1}_{\{t\geqslant T_k\}}e^{-\lambda_i T_k}e^{-(p-1)\lambda_i t}\eta_k^p. \tag{B.68}$$
Hence, (B.60), Lemma 3.2(ii), and (3.30) yield
$$\mathbb{E}\,|Z_1^*(t)|^p \leqslant C\,\mathbb{E}\sum_{k=1}^{\infty}\mathbb{E}\big[|Z_1^{(k)*}(t)|^p \mid T_k,\eta_k\big]$$
$$\leqslant Ce^{-(p-1)\lambda_i t}\,\mathbb{E}\sum_{k=1}^{\infty}\mathbf{1}_{\{t\geqslant T_k\}}e^{-\lambda_i T_k}\eta_k^p$$
$$= Ce^{-(p-1)\lambda_i t}\,\mathbb{E}\int_0^t e^{-\lambda_i s}X_j(s)\,\mathrm{d}s$$

$$\leqslant Ce^{-(p-1)\lambda_i t}\,\mathbb{E}\,\widetilde{X}_j^{**}\int_0^t (s+1)^{\kappa_j} e^{(\lambda_j^*-\lambda_i)s}\,\mathrm{d}s$$
$$\leqslant C(t+1)^{\kappa_j} e^{(\lambda_j^*-p\lambda_i)t}. \tag{B.69}$$

With $\widetilde{Z}_\ell^\dagger(n)$ defined in (3.70), we again modify (3.71) using pth powers, now using (B.69) instead of (3.68), and it follows similarly to (3.72) that

$$\mathbb{E}\sum_{n=1}^{\infty}\widetilde{Z}_\ell^\dagger(n)^p < \infty. \tag{B.70}$$

Hence, $\widetilde{Z}_1 \xrightarrow{\text{a.s.}} 0$.

Finally, in all cases, it follows from (3.43) that $\widetilde{X}_i(t) \xrightarrow{\text{a.s.}} \mathcal{X}_i$. Furthermore, we have $\mathcal{X}_j > 0 \implies \mathcal{X}_i > 0$ a.s. by the argument in Step 6 of Lemma 3.5, if necessary modified as in the proof of Lemma 8.11. □

(iii) *In Lemma 8.13,* $e^{-\delta t}X_i(t) \xrightarrow{\text{a.s.}} 0$ *still holds.*

Proof. The same as for Lemma 8.13. □

(iv) *In Lemmas 3.7 and 8.14,* $\widetilde{X}_i \xrightarrow{\text{a.s.}} \mathcal{X}_i$ *still holds, and so do the claims on* $\mathcal{X}_i > 0$.

Proof. The same as for Lemmas 3.7 and 8.14, using the claims above. Note that the assumptions now include (A7). □

To complete the proof of Theorem B.1, we now obtain by induction from (iv) above that for every $i \in \mathsf{Q}$ we have $\widetilde{X}_i \to \mathcal{X}_i$ as $t \to \infty$, i.e., (4.1). This proves that Theorems 4.1 and 8.6 hold with (A4) replaced by (A4p), which as said above completes the proof. □

Appendix C. Proof of (14.80)

In this appendix we prove (14.80) in Example 14.14, using results from Markov process theory.

Suppose, more generally, that the Pólya urn in Example 14.14 starts with $w_0 = \alpha > 0$ white balls and $b_0 \geqslant 0$ black balls. Thus $W(t) = \alpha$ for all $t \geqslant 0$. (We do not have to assume that α is an integer, although we must have $b_0 \in \mathbb{Z}_{\geqslant 0}$, since we allow subtractions.) The stochastic process $B(t)$ is a time-homogeneous pure-jump Markov process on $\mathbb{Z}_{\geqslant 0}$ with jumps

$$\begin{cases} +1 & \text{with intensity} \quad \alpha + \frac{1}{2}B(t), \\ -1 & \text{with intensity} \quad \frac{1}{2}B(t). \end{cases} \tag{C.1}$$

We define for any real $\ell > 0$ the scaled process

$$\widetilde{B}_\ell(t) := \ell^{-1}B(\ell t). \tag{C.2}$$

It follows from (C.1) that $\widetilde{B}_\ell$ is a pure-jump Markov process with jumps

$$\begin{cases} +1/\ell & \text{with intensity} \quad \ell\big(\alpha + \frac{1}{2}B(\ell t)\big) = \alpha\ell + \frac{\ell^2}{2}\widetilde{B}_\ell(t), \\ -1/\ell & \text{with intensity} \quad \frac{\ell^2}{2}\widetilde{B}_\ell(t). \end{cases} \tag{C.3}$$

In other words, the generator $\mathcal{A}_\ell$ of the Markov process $\widetilde{B}_\ell$ is given by, see e.g. [32, Theorem 19.23],

$$\mathcal{A}_\ell f(x) = \alpha\ell\big(f(x+\ell^{-1}) - f(x)\big) + x\frac{\ell^2}{2}\big(f(x+\ell^{-1}) + f(x-\ell^{-1}) - 2f(x)\big). \quad (C.4)$$

Since $B(t)$ takes its values in $\mathbb{Z}_{\geqslant 0}$, $\widetilde{B}_\ell(t)$ is a Markov process on the state space $\ell^{-1}\mathbb{Z}_{\geqslant 0} = \{0, \ell^{-1}, 2\ell^{-1}, \dots\}$. For technical reasons, we extend it to a pure-jump Markov process on $[0,\infty)$ by defining the intensities to be as in (C.3) whenever $\widetilde{B}_\ell(t) \geqslant \ell^{-1}$; otherwise, if $\widetilde{B}_\ell(t) = b < 1/\ell$, we jump $+1/\ell$ with intensity $\alpha\ell$ and $\pm b$ with intensities $\frac{\ell^2}{2}b$ each. (This makes no difference if we start with an integer number of black balls.) The generator of the extended process is

$$\mathcal{A}_\ell f(x) = \alpha\ell\big(f(x+\ell^{-1}) - f(x)\big) + x\frac{\ell^2}{2}\big(f(x+h_{\ell,x}) + f(x-h_{\ell,x}) - 2f(x)\big), \quad (C.5)$$

where $h_{\ell,x} := \ell^{-1} \wedge x$.

Let $C^2[0,\infty)$ be the space of all continuous functions f on $[0,\infty)$ that have two continuous derivatives in $(0,\infty)$ with f' and f'' extending continuously to $[0,\infty)$. Let further

$$\mathcal{C} := \big\{f \in C^2[0,\infty) : f(x), f'(x), f''(x) = O(e^{-\varepsilon x}) \text{ for some } \varepsilon > 0\big\}. \quad (C.6)$$

It follows from (C.5) and Taylor's formula that if $f \in \mathcal{C}$, then as $\ell \to \infty$,

$$\mathcal{A}_\ell f(x) \to \mathcal{A}f(x) := \alpha f'(x) + \frac{x}{2}f''(x) \quad (C.7)$$

uniformly in $x \in [0,\infty)$, and thus in the space $C_0[0,\infty)$. Note that $4\mathcal{A} = 2x\frac{\mathrm{d}^2}{\mathrm{d}x^2} + 4\alpha\frac{\mathrm{d}}{\mathrm{d}x}$ is the generator of the squared Bessel process BESQ^δ with dimension $\delta := 4\alpha$, see [44, p. 443 and Proposition VII.(1.7)]; hence $\mathcal{A}$ is the generator of $\mathrm{BESQ}^{4\alpha}(t/4) \overset{\mathrm{d}}{=} \frac{1}{4}\mathrm{BESQ}^{4\alpha}(t)$ [44, Proposition XI.(1.6)]; It is well known that $\mathrm{BESQ}^{4\alpha}(t)$ is a Feller process on $[0,\infty)$ [44, p. 442], and it is easily verified directly that each $\mathcal{A}_\ell$ also is a Feller process on $[0,\infty)$. The transition probabilities $q_t^\delta(x,y)$ of BESQ^δ are given explicitly in [44, XI.(1.4)], and a simple calculation using [40, (10.29.4)] shows that

$$\frac{\partial}{\partial x}q_t^\delta(x,y) = \frac{1}{2t}\big(q_t^{\delta+2}(x,y) - q_t^\delta(x,y)\big), \quad (C.8)$$

and hence

$$\frac{\partial^2}{\partial x^2}q_t^\delta(x,y) = \frac{1}{4t^2}\big(q_t^{\delta+4}(x,y) - 2q_t^{\delta+2}(x,y) + q_t^\delta(x,y)\big). \quad (C.9)$$

Since BESQ^δ is a Feller process for every $\delta > 0$, the transition operator $T_t^\delta f(x) := \int q_t^\delta(x,y)f(y)\,\mathrm{d}y$ maps $C_0[0,\infty)$ into itself for every $t > 0$, and it follows from (C.8)–(C.9) that T_t^δ maps $C_0[0,\infty)$ into $C^2[0,\infty)$; moreover, using also again the explicit form of q_t^δ in [44, XI.(1.4)], it is easy to see that T_t maps $\mathcal{C}$ into itself. Hence, it follows from [32, Proposition 19.9] that $\mathcal{C}$ is a core for the generator $4\mathcal{A}$, and thus also for $\mathcal{A}$. Consequently, (C.7) and [32, Theorem 19.25] show the following.

Theorem C.1. *Let the Pólya urn in Example 14.14 start with $w_0 = \alpha > 0$ white balls and $b_0 = 0$ black balls. Then, as $\ell \to \infty$, we have*

$$\ell^{-1}B(\ell t) = \widetilde{B}_\ell(t) \overset{\mathrm{d}}{\longrightarrow} \frac{1}{4}\mathrm{BESQ}^{4\alpha}(t) \qquad \text{in } D[0,\infty). \quad (C.10)$$

The squared Bessel process $\mathrm{BESQ}^{4\alpha}$ in (C.10) is the standard one starting at 0. More generally, if we start the urn with α white balls and $\beta\ell + o(\ell)$ black balls, then the same proof shows that (C.10) holds with the initial value $\frac{1}{4}\mathrm{BESQ}^{4\alpha}(0) = \beta$.

References

[1] Rafik Aguech. Limit theorems for random triangular urn schemes. *J. Appl. Probab.* **46** (2009), no. 3, 827–843. MR 2562324

[2] Krishna B. Athreya. On a characteristic property of Polya's urn. *Studia Sci. Math. Hungar.* **4** (1969), 31–35. MR 0247643

[3] Krishna B. Athreya & Samuel Karlin. Embedding of urn schemes into continuous time Markov branching processes and related limit theorems. *Ann. Math. Statist.* **39** (1968), 1801–1817. MR 0232455

[4] Krishna B. Athreya & Peter E. Ney. *Branching Processes.* Springer-Verlag, Berlin, 1972. MR 0373040

[5] Erich Baur. On a class of random walks with reinforced memory. *J. Stat. Phys.* **181** (2020), no. 3, 772–802. MR 4160910

[6] Erich Baur & Jean Bertoin. The fragmentation process of an infinite recursive tree and Ornstein–Uhlenbeck type processes. *Electron. J. Probab.* **20** (2015), no. 98, 20 pp. MR 3399834

[7] Erich Baur & Jean Bertoin. Elephant random walks and their connection to Pólya-type urns. *Physical Review* **94**, 052134 (2016).

[8] Bernard Bercu. On the elephant random walk with stops playing hide and seek with the Mittag-Leffler distribution. *J. Stat. Phys.* **189** (2022), no. 1, Paper No. 12, 27 pp. MR 4470258

[9] Nicholas H. Bingham. Continuous branching processes and spectral positivity. *Stochastic Process. Appl.* **4** (1976), no. 3, 217–242. MR 0410961

[10] David Blackwell & David Kendall. The Martin boundary of Pólya's urn scheme, and an application to stochastic population growth. *J. Appl. Probability* **1** (1964), 284–296. MR 0176518

[11] Arup Bose, Amites Dasgupta & Krishanu Maulik. Multicolor urn models with with reducible replacement matrices. *Bernoulli* **15** (2009), no. 1, 279–295. MR 2546808

[12] Arup Bose, Amites Dasgupta & Krishanu Maulik. Strong laws for balanced triangular urns. *J. Appl. Probab.* **46** (2009), no. 2, 571–584. MR 2535833

[13] Chen Chen & Hosam Mahmoud. The continuous-time triangular Pólya process. *Ann. Inst. Statist. Math.* **70** (2018), no. 2, 303–321. MR 3768935

[14] Colin Desmarais, Cecilia Holmgren & Stephan Wagner. Broadcasting-induced colorings of preferential attachment trees. *Random Structures Algorithms* **63** (2023), no. 2, 364–405. MR 4622330

[15] Michael Drmota, *Random Trees.* Springer-Verlag, Vienna, 2009. MR 2484382

[16] F. Eggenberger & George Pólya. Über die Statistik verketteter Vorgänge. *Zeitschrift Angew. Math. Mech.* **3** (1923), no. 4, 279–289.

[17] Philippe Flajolet, Philippe Dumas & Vincent Puyhaubert. Some exactly solvable models of urn process theory. *Proc. Fourth Colloquium on Mathematics and Computer Science Algorithms, Trees, Combinatorics and Probabilities*, Discrete Math. Theor. Comput. Sci. Proc. **AG**, 2006, 59–118. MR 2509623

[18] Raúl Gouet. A martingale approach to strong convergence in a generalized Pólya-Eggenberger urn model. *Statist. Probab. Lett.* **8** (1989), no. 3, 225–228. MR 1024032

[19] Raúl Gouet. Martingale functional central limit theorems for a generalized Pólya urn. *Ann. Probab.* **21** (1993), no. 3, 1624–1639. MR 1235432

[20] D. R. Grey. Asymptotic behaviour of continuous time, continuous state-space branching processes. *J. Appl. Probability* **11** (1974), 669–677. MR 0408016

[21] Allan Gut. *Probability: A Graduate Course*, 2nd ed., Springer, New York, 2013. MR 2977961

[22] Allan Gut & Ulrich Stadtmüller. The number of zeros in elephant random walks with delays. *Statist. Probab. Lett.* **174** (2021), Paper No. 109112, 9 pp. MR 4246204

[23] Allan Gut & Ulrich Stadtmüller. Elephant random walks with delays. *Rev. Roumaine Math. Pures Appl.* **67** (2022), no. 1-2, 51–66. MR 4429194

[24] Allan Gut & Ulrich Stadtmüller. Elephant random walks; a review. *Ann. Univ. Sci. Budapest. Sect. Comput.* **54** (2023), 171–198. MR 4660319

[25] Allan Gut & Ulrich Stadtmüller. Bernoulli elephant random walks. Preprint, 2022.

[26] Upendra Harbola, Niraj Kumar & Katja Lindenberg. Memory-induced anomalous dynamics in a minimal random walk model. *Physical Review E* **90** (2014), 022136.

[27] Svante Janson. Functional limit theorems for multitype branching processes and generalized Pólya urns. *Stoch. Process. Appl.* **110** (2004), 177–245. MR 2040966

[28] Svante Janson. Limit theorems for triangular urn schemes. *Probab. Theory Rel. Fields* **134** (2005), 417–452. MR 2226887

[29] Svante Janson. Rate of convergence for traditional Pólya urns. *J. Appl. Probab.* **57** (2020), no. 4, 1029-1044. MR 4179597

[30] Miloslav Jiřina. Stochastic branching processes with continuous state space. *Czechoslovak Math. J.* **8**:2 (1958), 292–313. MR 0101554

[31] Norman L. Johnson & Samuel Kotz. *Urn Models and Their Application.* John Wiley & Sons, New York, 1977. MR 0488211

[32] Olav Kallenberg. *Foundations of Modern Probability.* 2nd ed., Springer, New York, 2002. MR 1876169

[33] Niraj Kumar, Upendra Harbola & Katja Lindenberg. Memory-induced anomalous dynamics: Emergence of diffusion, subdiffusion, and superdiffusion from a single random walk model. *Physical Review E* **82** (2010), 021101.

[34] Andrea Kuntschik & Ralph Neininger. Rates of convergence for balanced irreducible two-color Pólya urns. *2017 Proceedings of the Fourteenth Workshop on Analytic Algorithmics and Combinatorics (ANALCO)*, 94-–99, SIAM, Philadelphia, PA, 2017. MR 3630937

[35] Rüdiger Kürsten. Random recursive trees and the elephant random walk. *Physical Review E* **93** (2016), 032111.

[36] John Lamperti. Continuous state branching processes. *Bull. Amer. Math. Soc.* **73** (1967), 382–386. MR 0208685

[37] Zenghu Li. *Measure-Valued Branching Markov Processes.* Springer, Berlin, Heidelberg, 2011. MR 2760602

[38] Hosam M. Mahmoud. *Pólya Urn Models.* CRC Press, Boca Raton, FL, 2009. MR 2435823

[39] Andrey A. Markov. Sur quelques formules limites du calcul des probabilités. (Russian.) *Bulletin de l'Académie Impériale des Sciences* **11** (1917), no. 3, 177–186.

[40] *NIST Handbook of Mathematical Functions.* Edited by Frank W. J. Olver, Daniel W. Lozier, Ronald F. Boisvert and Charles W. Clark. Cambridge Univ. Press, 2010.
Also available as *NIST Digital Library of Mathematical Functions*, `http://dlmf.nist.gov/` MR 2723248

[41] George Pólya. Sur quelques points de la théorie des probabilités. *Ann. Inst. Poincaré* **1** (1930), no. 2, 117–161. MR 1507985

[42] Philip E. Protter. *Stochastic Integration and Differential Equations*, 2nd ed. Springer-Verlag, Berlin, 2004. MR 2020294

[43] Vincent Puyhaubert. *Modèles d'urnes et phénomènes de seuils en combinatoire analytique*. Ph. D. thesis, l'École Polytechnique, 2005. Available at `https://pastel.hal.science/pastel-00001348`

[44] Daniel Revuz & Marc Yor. *Continuous Martingales and Brownian Motion.* 3rd ed., Springer-Verlag, Berlin, 1999. MR 1725357.

Department of Mathematics, Uppsala University, PO Box 480, SE-751 06 Uppsala, Sweden

Email address: `svante.janson@math.uu.se`

URL: `http://www.math.uu.se/~svante/`