FUNCTIONAL LIMIT THEOREMS FOR MULTITYPE BRANCHING PROCESSES AND GENERALIZED PÓLYA URNS

SVANTE JANSON

Abstract. A functional limit theorem is proved for multitype continuous time Markov branching processes. As consequences, we obtain limit theorems for the branching process stopped by some stopping rule, for example when the total number of particles reaches a given level.

Using the Athreya–Karlin embedding, these results yield asymptotic results for generalized Pólya urns. We investigate such results in detail and obtain explicit formulas for the asymptotic variances and covariances. The general formulas involve integrals of matrix functions; we show how they can be evaluated and simplified in important special cases. We also consider the numbers of drawn balls of different types and functional limit theorems for the urns.

We illustrate our results by some examples, including several applications to random trees where our theorems and variance formulas give simple proofs of some known results; we also give some new results.

1. Introduction

Consider a generalized Pólya urn process $(X_n)_{n=0}^{\infty}$ defined as follows. (This process is also known as a generalized Pólya–Eggenberger urn or a generalized Friedman urn, cf. [17], [47], [19].) There are balls of q types (or colours) $1, \dots, q$, and each X_n is a vector $(X_{n1}, \dots, X_{nq})$, where $X_{ni} \geq 0$ is the number of balls of type i in the urn at time n. The urn starts with a given vector X_0, random or not. We are further given, for each type i, an activity (or weight) $a_i \geq 0$ (typically $a_i = 1$, but sometimes different a_i are useful [2]; we will even find use for $a_i = 0$), and a random q-dimensional vector $\xi_i = (\xi_{i1}, \dots, \xi_{iq})$ with integer coordinates. (Actually, only the distribution of ξ_i matters.) We usually further assume that, almost surely,

$$\xi_{ij} \geq 0, \qquad j \neq i, \tag{1.1}$$

$$\xi_{ii} \geq -1. \tag{1.2}$$

(Relaxation of these requirements, allowing further negative values, will be discussed in Remark 4.2.) The urn evolves according to a Markov process. At each time $n \geq 1$, one of the balls in the urn is drawn at random such that the probability of drawing a particular ball of type i is proportional to the activity a_i, i.e. the probability of drawing a ball of type i

Date: February 20, 2003.

is $a_i X_{n-1,i}/\sum_j a_j X_{n-1,j}$. (In particular, if every $a_i = 1$, a ball is drawn uniformly at random.) The drawn ball is returned to the urn together with ΔX_{nj} balls of type j, for each $j = 1, \dots, q$, where $\Delta X_n = (\Delta X_{n1}, \dots, \Delta X_{nq})$ is a random vector such that if the drawn ball has type i, then ΔX_n has the same distribution as ξ_i and is independent of everything else that has happened so far. (In many applications, the replacement vectors ξ_i are deterministic and the randomness enters solely through the the random draws.)

Note that (1.1) means that we may add but never remove balls of other types than the drawn one, while (1.2) means that $\Delta X_{ni} = -1$ is allowed when i is drawn, meaning that the drawn ball is removed (with or without addition of balls of other types). Indeed, the rule above may also be stated as: The drawn ball is removed and, if it had type i, we add a number of new balls with the distribution of $(\xi_{ij} + \delta_{ij})_{j=1}^q$. This is often a more natural formulation, and explains (1.2) better. Note that, in both versions, ξ_i describes the *change* in the composition of the urn when a ball of type i is drawn.

If the urn ever becomes empty, or, more generally, there are no balls with non-zero activity left, the process stops (*extinction*). We are only interested in urns where there is a positive probability of non-extinction, and our main goal is to describe the asymptotics of the urn conditioned on non-extinction. Indeed, in typical applications extinction cannot occur at all.

For some specific examples and applications, see Section 7.

Urn models of this type have been studied by many authors, including [13], [14], [49], [19], [18], [8], [9], [30], [10], [20], [51], [11], [35], [12].

We will use the method of Athreya and Karlin [8], see also [9, §V.9], and study the urn process by embedding it into a multitype continuous time Markov branching process $\mathcal{X}(t) = \big(\mathcal{X}_1(t), \dots, \mathcal{X}_q(t)\big)$. This process is defined using the same data a_i and ξ_i, $i = 1, \dots, q$, as above and an initial vector $\mathcal{X}(0) = X_0$. In this process we assume that a ball (particle) of type i lives an exponentially distributed time with mean a_i^{-1}, i.e. it dies with intensity a_i, and when it dies, it is replaced by a set of balls with the distribution given by $(\xi_{ij} + \delta_{ij})_{j=1}^q$, all life times and offspring compositions being independent. Alternatively, when $\xi_{ii} \ge 0$ a.s., the ball lives for ever and at random times according to a Poisson process with intensity a_i, it gives birth to a new litter of balls with the distribution given by ξ_i. (Unfortunately, the embedding is exemplified in [8] and [9] only with urn processes where the drawn ball is returned, which in our notation means $\xi_{ii} \ge 0$. It is clear from [8], [9] that the results hold also when the drawn ball is removed, as observed in e.g. [10], [2], but some authors have overlooked this.) We assume throughout that all variables ξ_i have finite mean (and variance, see (A2) below); this is sufficient to prevent explosion and guarantees that the process $\mathcal{X}(t)$ exists for all $t \ge 0$ [9, §V.7]. We define $\mathcal{X}(t)$ to be right-continuous.

Let $\tau_0 = 0$, and let τ_n, $n \ge 1$, be the n:th time a ball dies (splits). It is easily shown [8], [9] that the process $(\mathcal{X}(\tau_n))_{n=0}^\infty$ equals (in distribution)

$(X_n)_{n=0}^{\infty}$; hence limit theorems for X_n can be derived from limit theorems for $\mathcal{X}(t)$. The processes $\mathcal{X}(t)$, $t \geq 0$, and X_n, $n \geq 0$, are the same up to a random time change (extending the parameter n to real values). However, since $\mathcal{X}(t)$ grows exponentially (Lemma 9.8), the time scales are different.

In several applications, see e.g. Examples 7.5 and 7.8, urn processes not satisfying (1.1) and (1.2) appear. Although the embedding method by Athreya and Karlin does not apply directly, it can be modified to handle this case too, in at least two ways. First, in many cases, it is possible to transform the urn process into a different urn process satisfying the conditions above, using a "superball" argument, see Remark 4.2. Secondly, even if ξ_{ij} are arbitrary integers, but we for simplicity assume that they are such that the process never gets stuck, it is possible to define the corresponding continuous time process; this is a generalized branching process where the death of one ball may force the removal of others (like the ancient practices of sacrificing slaves or burning widows), and under appropriate conditions, the results extend to this case too. For simplicity, we nevertheless assume (1.1) and (1.2) in the main parts of the paper, and discuss these extensions in Remark 4.2 and some examples.

Our main results are stated in Section 3; we first introduce some notation and basic assumptions in Section 2. Some extensions, and problems for future research, are discussed in Section 4. We aim at directly applicable results where, in the case of normal limits, the asymptotic variances and covariances are given explicitly, by formulas computable using linear algebra. The formulas in Section 3 are given by integrals. Some simplifications and evaluations of the formulas in important special cases are given in Section 5; see also the examples in Section 7. We discuss how our results and methods relate to some previous papers in Section 6.

In Section 7, we give some examples and applications of our results. In particular, urn processes have been used by several authors to study various classes of random trees, see e.g. the "fringe analysis" in [2]. We review several such applications and show how some old (and a few new) results follow easily from our theorems. See also [29], which may serve as an easier introduction to such applications. Finally, the proofs of the main results are given in Sections 8–10.

Acknowledgements. This research was begun more than ten years ago, inspired by a talk by Hosam Mahmoud at a conference on Random Graphs in Poznań 1991; I thank Hosam Mahmoud and others for interesting discussions over the years. The research has partly been carried out during visits to the Mittag-Leffler Institute in Djursholm, Sweden and the Isaac Newton Institute for Mathematical Sciences in Cambridge, U.K.

2. Preliminaries

We let A denote the $q \times q$ matrix

$$A := (a_j \, \mathbb{E}\, \xi_{ji})_{i,j=1}^{q}. \tag{2.1}$$

The matrix A and its eigenvalues will play a central role.

Note our choice of notation; in the main case when $a_j = 1$, $A_{ij} = \mathbb{E}\,\xi_{ji}$ and the j:th *column* of A is the expected change when a ball of type j is drawn (splits). It may seem more natural to consider the transpose A', and this is done by other authors (which should be remembered when comparing the results). The reason for our choice is that we will use the standard notation where a matrix is regarded as an operator acting on column vectors to the right. (In contrast to the standard notation for Markov chains, where the transition matrix acts on row vectors to the left.)

By (1.1), $A + \alpha I$ is a non-negative matrix if α is large enough, so by the standard Perron–Frobenius theory, A has a largest real eigenvalue λ_1 such that every other eigenvalue λ satisfies $\operatorname{Re}\lambda < \lambda_1$ (see e.g. [50, Chapter 1 and Theorem 2.6] or [32, Appendix 2]). We order the eigenvalues with decreasing real parts: $\lambda_1 > \operatorname{Re}\lambda_2 \geq \operatorname{Re}\lambda_3 \geq \dots$.

We write $i \succ j$ if it is possible to find a ball of type j in an urn beginning with a single ball of type i, i.e. if $(A^n)_{ji} > 0$ for some $n \geq 0$. The relation $\succ$ is transitive and reflexive, so it partitions the set of all types into equivalence classes $\mathcal{C}_1, \dots, \mathcal{C}_\nu$ such that i and j belong to the same class if and only if $i \succ j$ and $j \succ i$; moreover, $\succ$ induces a partial order among the equivalence classes. We say that a type i is *dominating* if $i \succ j$ for every type j; similarly a class $\mathcal{C}_k$ is dominating if some (and then every) $i \in \mathcal{C}_k$ is dominating.

Note that if we order the classes suitably, and take the types in this order, A becomes a block triangular matrix, see [34] for a detailed treatment. Hence the set of eigenvalues of A (with multiplicities) is the union of the sets of eigenvalues of the restrictions of A to the classes $\mathcal{C}_k$; we say that an eigenvalue belongs to a class if it is an eigenvalue of the restriction of A to this class.

The urn or branching process (or A) is *irreducible* (or *positive regular*, which is equivalent in continuous time) if there is only one equivalence class, i.e. if $i \succ j$ for any types i and j; equivalently, every type is dominating.

We are mainly interested in the irreducible case, but for an important technical reason, see for example the proof of Theorem 3.16, we will state our results somewhat more generally.

Our basic assumptions are as follows (but see Remark 4.2):

- (A1) (1.1) and (1.2) hold, i.e. $\xi_{ij} + \delta_{ij} \geq 0$ a.s. for all i and j.
- (A2) $\mathbb{E}\,\xi_{ij}^2 < \infty$ for all $i, j = 1, \dots, q$.
- (A3) The largest real eigenvalue λ_1 of A is positive, $\lambda_1 > 0$.
- (A4) The largest real eigenvalue λ_1 is simple.
- (A5) There exists a dominating type i with $X_{0i} > 0$ ($\mathcal{X}(0)_i > 0$), i.e. we start with at least one ball of a dominating type.
- (A6) λ_1 belongs to the dominating class.

We assume that the classes are ordered so that $\mathcal{C}_1$ is the dominating class.

(A1) is already discussed. (A2) is essential since we use L^2 theory. (A3) says that the branching process $\mathcal{X}(t)$ is supercritical, and implies that $\mathcal{X}(t)$, and thus X_n, has a positive probability of non-extinction. (Non-extinction is also possible in some exceptional cases with $\lambda_1 = 0$, for example when the balls always have exactly one child, and thus change type according to a Markov chain. We do not treat these cases.)

Note that (A4), (A5) and (A6) hold when A is irreducible [32], [50]. In the reducible case, (A5) is only a weak restriction; if we consider a case with a single ball initially, we may ignore all types that cannot occur and then (A5) holds. (A4) and (A6) are more significant restrictions; see [34] (which treats the related case of multitype Galton–Watson processes in discrete time) or [35] for some complications that otherwise can occur.

We say that the process becomes *essentially extinct* if at some time there are no balls of any dominating type left. Note that if we restrict attention to the balls of the dominating types, we have an irreducible multitype Galton–Watson process, and essential extinction means that this restricted process becomes extinct. For irreducible processes (our main concern), essential extinction is thus the same as extinction.

For most applications, the following lemma yields convenient criteria (possibly combined with Remark 4.2).

Lemma 2.1. *If A is irreducible,* (A1) *and* (A2) *hold,* $\sum_j \mathbb{E}\,\xi_{ij} \geq 0$ *for every i and* $\sum_j \mathbb{E}\,\xi_{ij} > 0$ *for some i, then* (A1)–(A6) *hold and (essential) extinction is impossible.*

Proof. The conditions imply that the total number of balls never decreases, which guarantees non-extinction. Since the process is irreducible, this is the same as essential non-extinction. We have already remarked that (A4)–(A6) hold when A is irreducible. Finally, it is easy to see that the conditions imply (A3), cf. [50, Theorem 1.1, Corollary 1]. □

We collect various facts and notations that will be used throughout the paper, usually without further comment.

We will often regard A and other matrices as linear operators in $\mathbb{R}^q$ or $\mathbb{C}^q$, or in some invariant subspace thereof. In this context, vectors in $\mathbb{C}^q$, in particular X_n and $\mathcal{X}(t)$, are always regarded as column vectors. Consequently, by an eigenvector of A we mean a right eigenvector; a left eigenvector is the same as an eigenvector of the transpose matrix A'.

Note that if u and v are vectors in $\mathbb{C}^q$, then $u'v$ is a scalar while uv' is a $q \times q$ matrix. We also use the notation $u \cdot v$ for $u'v$.

We use $|\cdot|$ for the norm of both vectors and matrices. (The choice of matrix norm is irrelevant.)

We let a denote the (column) vector $(a_1, \dots, a_q)$ of activities, and let u_1 and v_1 denote left and right eigenvectors of A corresponding to the largest eigenvalue λ_1, i.e. vectors satisfying

$$u_1' A = \lambda_1 u_1', \qquad A v_1 = \lambda_1 v_1.$$

By (A4) u_1 and v_1 are unique up to scalar factors, and by the Perron–Frobenius theory [32], [50], (applied to $A + \alpha I$ for suitable α), they may be chosen non-negative.

If the process is irreducible, all entries of u_1 and v_1 are strictly positive [32], [50]. In general, it follows easily from this result applied to the restriction to the dominating class $\mathcal{C}_1$ together with (A6) that $v_{1i} > 0$ for every i while $u_{1i} > 0$ if $i \in \mathcal{C}_1$ (i.e., i is dominating) and $u_{1i} = 0$ otherwise.

The scalar products $u_1 \cdot v_1$ and $a \cdot v_1$ thus are both positive, and we may assume that v_1 and u_1 are normalized such that

$$a \cdot v_1 = a' v_1 = v_1' a = 1, \tag{2.2}$$

$$u_1 \cdot v_1 = u_1' v_1 = v_1' u_1 = 1. \tag{2.3}$$

This determines u_1 and v_1, and we fix this choice of u_1 and v_1 throughout the paper. (Otherwise, obvious normalization factors would enter into some formulas.)

We will use the Jordan decomposition of the matrix A in the following form, see e.g. [46, Theorem 7.6]: There exists a decomposition of the *complex* space $\mathbb{C}^q$ as a direct sum $\bigoplus E_\lambda$ of generalized eigenspaces E_λ, such that $A - \lambda$ is a nilpotent operator on E_λ; here λ ranges over the set Λ of eigenvalues of A. ($A - \lambda$ denotes $A - \lambda I$, where I is the identity matrix of appropriate size.) In other words, there exist projections P_λ, $\lambda \in \Lambda$, that commute with A and satisfy

$$\sum_{\lambda \in \Lambda} P_\lambda = I, \tag{2.4}$$

$$A P_\lambda = P_\lambda A = \lambda P_\lambda + N_\lambda, \tag{2.5}$$

where $N_\lambda = P_\lambda N_\lambda = N_\lambda P_\lambda$ is nilpotent. Moreover, $P_\lambda P_\mu = 0$ when $\lambda \neq \mu$. We let $d_\lambda \geq 0$ be the integer such that $N_\lambda^{d_\lambda} \neq 0$ but $N_\lambda^{d_\lambda + 1} = 0$. (Equivalently, in the Jordan normal form of A, the largest Jordan block with λ on the diagonal has size $d_\lambda + 1$.) Hence $d_\lambda = 0$ if and only if $N_\lambda = 0$, and this happens for all λ if and only if A is diagonalizable, i.e. if and only if A has a complete set of q linearly independent eigenvectors.

Note, by taking transposes in (2.4) and (2.5), that P_λ' and N_λ' are the corresponding projections and nilpotent operators for A'.

We define, for $k = 0, 1, \dots$, the quotient space $E_{\lambda,k} := E_\lambda / N_\lambda^{k+1} E_\lambda$ and the projection $Q_{\lambda,k} : E_\lambda \to E_{\lambda,k}$, noting that $E_{\lambda,d_\lambda} = E_\lambda$ and $Q_{\lambda,d_\lambda} = I$. Then $N_\lambda : E_\lambda \to E_\lambda$ induces a map $N_\lambda : E_{\lambda,k} \to E_{\lambda,k+1}$, and if $0 \leq j \leq k$,

$$N_\lambda^j Q_{\lambda,k-j} = Q_{\lambda,k} N_\lambda^j : E_\lambda \to E_{\lambda,k}. \tag{2.6}$$

Since we assume that λ_1 is a simple eigenvalue, $N_{\lambda_1} = 0$ and $d_{\lambda_1} = 0$, and P_{λ_1} is the one-dimensional projection

$$P_{\lambda_1} = v_1 u_1'. \tag{2.7}$$

In the sequel, λ will always denote an eigenvalue of A. (Formally, the results hold for other λ too if we then set $P_\lambda = N_\lambda = 0$.)

We recall that matrix exponentials may be defined by power series; for example $e^{tA} = \sum_{j=0}^{\infty} t^j A^j / j!$. We have, using (2.5) and commutativity,

$$\begin{aligned} P_\lambda e^{tA} = e^{tA} P_\lambda = P_\lambda \sum_{j=0}^{\infty} \frac{t^j}{j!} (P_\lambda A)^j = P_\lambda e^{tP_\lambda A} = P_\lambda e^{\lambda t P_\lambda + tN_\lambda} \\ = P_\lambda e^{\lambda t P_\lambda} e^{tN_\lambda} = P_\lambda e^{\lambda t} \sum_{j=0}^{d_\lambda} \frac{t^j}{j!} N_\lambda^j \end{aligned} \tag{2.8}$$

and thus, by (2.4),

$$e^{tA} = \sum_\lambda \sum_{j=0}^{d_\lambda} \frac{t^j}{j!} e^{\lambda t} P_\lambda N_\lambda^j. \tag{2.9}$$

Some immediate consequences are

$$|P_\lambda e^{tA}| \le C(1+|t|)^{d_\lambda} e^{\operatorname{Re}\lambda t}, \qquad -\infty < t < \infty, \tag{2.10}$$

and more generally, for $0 \le k \le d_\lambda$,

$$|Q_{\lambda,k} P_\lambda e^{tA}| \le C(1+|t|)^{k} e^{\operatorname{Re}\lambda t}, \qquad -\infty < t < \infty, \tag{2.11}$$

and, using (A4),

$$|e^{tA}| \le C e^{\lambda_1 t}, \qquad 0 \le t < \infty, \tag{2.12}$$

where, as sometimes later, C denotes unspecified constants that may depend on the data q, a_i, ξ_i, X_0.

As is well-known since decades [33], [5], [6], the asymptotic behavior depends on whether there is any eigenvalue beside λ_1 with a real part $> \lambda_1/2$. We thus define $\Lambda_I := \{\lambda \in \Lambda : \operatorname{Re}\lambda < \lambda_1/2\}$, $\Lambda_{II} := \{\lambda \in \Lambda : \operatorname{Re}\lambda = \lambda_1/2\}$, $\Lambda_{III} := \{\lambda \in \Lambda : \operatorname{Re}\lambda > \lambda_1/2\}$; hence Λ is the disjoint union $\Lambda_I \cup \Lambda_{II} \cup \Lambda_{III}$. We further define $P_I := \sum_{\lambda \in \Lambda_I} P_\lambda$, the projection onto the sum of the generalized eigenspaces with $\operatorname{Re}\lambda < \lambda_1/2$.

For later use, we define the following matrices.

$$B_i := \mathbb{E}(\xi_i \xi_i'), \tag{2.13}$$

$$B := \sum_{i=1}^{q} v_{1i} a_i B_i, \tag{2.14}$$

$$\Sigma_I := \int_0^\infty P_I e^{sA} B e^{sA'} P_I' e^{-\lambda_1 s}\, ds, \tag{2.15}$$

$$\Sigma_{II} := \sum_{\lambda \in \Lambda_{II}} P_\lambda B P_\lambda' = \sum_{\lambda \in \Lambda_{II}} P_\lambda B P_\lambda^*, \tag{2.16}$$

$$\Sigma_{II,d} := \frac{1}{(2d+1)d!^2} \sum_{\lambda \in \Lambda_{II}} N_\lambda^d P_\lambda B P_\lambda^* (N_\lambda^*)^d, \tag{2.17}$$

where * denotes Hermite conjugation and $d = 0, 1, \dots$; thus $\Sigma_{II,0} = \Sigma_{II}$.

Remark 2.2. Let ξ_* be the random vector obtained by choosing ξ_i for a random type i, with the probability $a_i v_{1i}$ for type i. It follows from Theorem 3.21 below that this is the asymptotic distribution of the drawn types; thus ξ_* is the asymptotic distribution of the added balls.

Then (2.14) says $B = \mathbb{E}\,\xi_*\xi_*'$. Since, $\mathbb{E}\,\xi_* = \sum_{i=1}^q v_{1i}a_i\,\mathbb{E}\,\xi_i$ and, by (2.1),

$$\sum_{i=1}^q v_{1i}a_i\,\mathbb{E}\,\xi_i = \Bigl(\sum_{i=1}^q v_{1i}a_i\,\mathbb{E}\,\xi_{ij}\Bigr)_{j=1}^q = \Bigl(\sum_{i=1}^q v_{1i}A_{ji}\Bigr)_{j=1}^q = Av_1 = \lambda_1 v_1, \tag{2.18}$$

the covariance matrix of ξ_* is given by

$$\widehat{B} := \mathbb{E}\,\xi_*\xi_*' - \mathbb{E}\,\xi_*\,\mathbb{E}\,\xi_*' = B - \lambda_1^2 v_1 v_1'.$$

Not surprisingly, this quantity will appear below. Indeed, since $P_\lambda v_1 = 0$ when $\lambda \neq \lambda_1$, we may replace B by $\widehat{B}$ in (2.15)–(2.17); this might be conceptually better, but we prefer B for computations.

We state our results on convergence of stochastic processes using the usual function space D of right-continuous functions with left-hand limits, always equipped with the Skorohod J_1-topology. Our processes will, however, be defined on several different intervals, so we will use several versions of D; we will also consider vector-valued processes. In general, for a finite-dimensional vector space E and any (open, closed or half-open) interval $J \subseteq [-\infty,\infty]$, we let $D(J) = D(J,E)$ be the space of all right-continuous functions $J \to E$ with left-hand limits. We say that $f_n \to f$ in D if there exists strictly increasing continuous maps λ_n of J onto itself such that $\lambda_n \to \iota$ (the identity map) and $f_n \circ \lambda_n \to f$ uniformly on compact subsets of J. When f is continuous, this is equivalent to $f_n \to f$ uniformly on compact subsets of J. It is well-known that this topology is Polish, i.e. defined by some complete metric. (The case $J = [0,1]$ is discussed in detail in [15], and the case $J = [0,\infty)$ in [37], [24]. See also [27].) Note that both the space and the topology are changed if we add or remove an endpoint of J. If $Z_n \overset{\mathrm{d}}{\to} Z$ in $D(J)$ for some processes Z_n and Z defined on J, and Z is a.s. continuous, then the restrictions to any subinterval $J' \subset J$ converge in $D(J')$.

3. Results

The basis of all our results for the branching process and generalized Pólya urns, is the following functional limit theorem for the branching process. (For previously known results, including parts of this theorem, see Section 6.) Since different normalizations (and different time scalings) are required for different components of $\mathcal{X}(t)$, the result is stated in terms of various projections of $\mathcal{X}(t)$; this is equivalent to stating results for scalar products $\eta \cdot \mathcal{X}(t)$, where the normalization depends on η, as is done by several other authors. The proof is given in Section 9.

Theorem 3.1. *Assume* (A1)–(A6). *Then, as* $t \to \infty$, $e^{-\lambda_1 t}\mathcal{X}(t) \overset{\text{a.s.}}{\to} W v_1$, *and, with joint convergence in distribution of all processes,*

(i) $e^{-\lambda_1(t+x)/2}P_I\mathcal{X}(t+x)\overset{\mathrm{d}}{\to}W^{1/2}U_I(x)$ *in* $D(-\infty,\infty)$; *equivalently,* $e^{-\lambda_1 t/2}P_I\mathcal{X}(t+x)\overset{\mathrm{d}}{\to}W^{1/2}e^{\lambda_1 x/2}U_I(x)$ *in* $D(-\infty,\infty)$;

(ii) *for every* $\lambda\in\Lambda_{II}$ *and* $k=0,\dots,d_\lambda$,
$$t^{-(k+1/2)}e^{-\lambda xt}Q_{\lambda,k}P_\lambda\mathcal{X}(xt)\overset{\mathrm{d}}{\to}W^{1/2}U_{\lambda,k}(x)\quad\text{in } D[0,\infty);$$

(iii) *for every* $\lambda\in\Lambda_{III}$ *and* $k=0,\dots,d_\lambda$, $t^{-k}e^{-\lambda t}Q_{\lambda,k}P_\lambda\mathcal{X}(t)\overset{\text{a.s.}}{\to}W_{\lambda,k}$ *and thus*
$$(xt)^{-k}e^{-\lambda xt}Q_{\lambda,k}P_\lambda\mathcal{X}(xt)\overset{\mathrm{d}}{\to}W_{\lambda,k}\quad\text{in } D(0,\infty).$$

Here, U_I *and* $U_{\lambda,k}$ *are continuous Gaussian vector-valued stochastic processes, with* $U_I(x)$ *defined for* $-\infty<x<\infty$ *and* $U_{\lambda,k}(x)$ *defined for* $0\le x<\infty$, *while* $W_{\lambda,k}$ *are vector-valued random variables, also regarded as constant stochastic processes. Moreover,* W *is a non-negative random variable, related to* $W_{\lambda_1}:=W_{\lambda_1,0}$ *by* $W=u_1\cdot W_{\lambda_1}$ *and* $W_{\lambda_1}=Wv_1$.

The process U_I *is real, while the processes* $U_{\lambda,k}$ *and variables* $W_{\lambda,k}$ *are real for real* λ *but complex otherwise, with* $U_{\bar\lambda,k}=\overline{U_{\lambda,k}}$ *and* $W_{\bar\lambda,k}=\overline{W_{\lambda,k}}$. *Furthermore, a.s.,* $U_I(x)\in E_I:=\bigoplus_{\lambda\in\Lambda_I}E_\lambda$, $U_{\lambda,k}(x)\in E_{\lambda,k}$, *and* $W_{\lambda,k}\in E_{\lambda,k}$.

The process U_I, *the families* $\{U_{\lambda,k}\}_{0\le k\le d_\lambda}$ *for different* $\lambda\in\Lambda_{II}$ *with* $\operatorname{Im}\lambda\ge 0$, *and the family* $\{W_{\lambda,k}\}_{\lambda\in\Lambda_{III},k\le d_\lambda}\cup\{W\}$ *are independent of each other.*

The processes U_I *and* $U_{\lambda,k}$ *are characterized by being (jointly) Gaussian with mean* 0 *and covariances, for* $0\le x\le y$,
$$\mathbb{E}\big(U_I(x)U_I(y)'\big)=\Sigma_I e^{(y-x)(A'-\lambda_1/2)},\tag{3.1}$$
and
$$\mathbb{E}\big(U_{\lambda,k}(x)U_{\mu,l}(y)'\big)=c(k,l,x,y)Q_{\lambda,k}N_\lambda^kP_\lambda\Sigma_{II}P_\mu'(N_\mu')^lQ_{\mu,l}'\tag{3.2}$$
(which vanishes unless $\mu=\bar\lambda$*), where*
$$\begin{aligned}c(k,l,x,y)&:=\frac{1}{k!\,l!}\int_0^x s^k(y-x+s)^l\,ds\\&=\frac{1}{k!\,l!}\sum_{j=0}^{l}\binom{l}{j}(y-x)^j\frac{x^{k+l+1-j}}{k+l+1-j}.\end{aligned}\tag{3.3}$$
The results above holds also if we condition $\mathcal{X}$ *and* W, $W_{\lambda,k}$ *on* $W>0$, *or (which is a.s. the same) on essential non-extinction.*

Remark 3.2. When $x=y$, (3.3) simplifies to
$$c(k,l,x,x)=\frac{x^{k+l+1}}{k!\,l!\,(k+l+1)}.$$

Remark 3.3. Taking $k=d_\lambda$ in (ii) or (iii), we have $Q_{\lambda,k}=I$ and we thus find the limit of $P_\lambda\mathcal{X}$ under appropriate normalization. The point of the variable k is that, when $d_\lambda>0$ (i.e., when the nilpotent part N_λ

does not vanish), some linear combinations of the components of $P_\lambda \mathcal{X}$ have smaller asymptotic variance than others, and thus they disappear in the normalization required for $P_\lambda \mathcal{X}$. More precisely, if $\eta \in E'_\lambda$ is such that $(N'_\lambda)^{k+1}\eta = 0$, then η can be regarded as a linear functional on $E_{\lambda,k}$, and if, for example, $\lambda \in \Lambda_{II}$, (ii) shows that $t^{-k-1/2}e^{-\lambda_1 t/2}\eta \cdot \mathcal{X}(xt)$ converges to a Gaussian process.

Remark 3.4. If $\lambda \in \Lambda_{II}$ and $d_\lambda = 0$, (ii) simplifies to $t^{-1/2}e^{-\lambda xt}P_\lambda \mathcal{X}(xt) \overset{\mathrm{d}}{\to} W^{1/2}U_\lambda(x)$, with, from (3.2) and (3.3), for $0 \le x \le y$,

$$\mathbb{E}\big(U_\lambda(x)U^*_\lambda(y)\big) = xP_\lambda \Sigma_{II} P^*_\lambda = xP_\lambda B P^*_\lambda.$$

In this case, U_λ is a process with independent increments.

Remark 3.5. By (3.1), U_I is a stationary Gaussian process. It can be regarded as a multi-dimensional Ornstein–Uhlenbeck process.

Remark 3.6. If $\lambda \in \Lambda_{II}$ with $\operatorname{Im}\lambda \neq 0$, it follows from (3.2) and (2.16) that $\mathbb{E}\big(U_{\lambda,k}(x)U_{\lambda,k}(x)'\big) = 0$. Hence $U_{\lambda,k}(x)$ is a vector-valued symmetric complex Gaussian random variable, i.e. $\omega U_{\lambda,k}(x) \overset{\mathrm{d}}{=} U_{\lambda,k}(x)$ for every complex number ω with $|\omega| = 1$, see [28, Proposition 1.31].

Consequently, $U_{\lambda,k}$ is either real (when λ is real) or symmetric complex.

Remark 3.7. We have no general description of the distributions of $W_{\lambda,k}$ for $\lambda \in \Lambda_{III}$, and there seems to be no reason to expect any. They are (typically, at least) not normal, and not independent of each other. Moreover, their distributions (typically) depend on the initial state $\mathcal{X}(0)$, unlike U_I and $U_{\lambda,k}$.

Taking $x = 0$ in (i) and $x = 1$ in (ii) and (iii), we obtain as a corollary some standard results, cf. [9].

Corollary 3.8. *Assume* (A1)–(A6). *Then, as $t \to \infty$, with joint convergence,*

(i) $e^{-\lambda_1 t/2}P_I \mathcal{X}(t) \overset{\mathrm{d}}{\to} W^{1/2}U_I$;
(ii) *for every $\lambda \in \Lambda_{II}$ and $k = 0, \dots, d_\lambda$,*

$$t^{-(k+1/2)}e^{-\lambda t}Q_{\lambda,k}P_\lambda \mathcal{X}(t) \overset{\mathrm{d}}{\to} W^{1/2}U_{\lambda,k};$$

(iii) *for every $\lambda \in \Lambda_{III}$ and $k = 0, \dots, d_\lambda$, $t^{-k}e^{-\lambda t}Q_{\lambda,k}P_\lambda \mathcal{X}(t) \overset{\mathrm{d}}{\to} W_{\lambda,k}$ (and $\overset{\mathrm{a.s.}}{\to}$).*

Here, U_I, $U_{\lambda,k}$ and $W_{\lambda,k}$ are vector-valued random variables with U_I and $U_{\lambda,k}$ jointly Gaussian. The vector U_I is real, while $U_{\lambda,k}$ and $W_{\lambda,k}$ are real for real λ but complex otherwise, with $U_{\bar\lambda,k} = \overline{U_{\lambda,k}}$ and $W_{\bar\lambda,k} = \overline{W_{\lambda,k}}$. Furthermore, a.s., $U_I \in E_I := \bigoplus_{\lambda\in\Lambda_I} E_\lambda$, $U_{\lambda,k} \in E_{\lambda,k}$, and $W_{\lambda,k} \in E_{\lambda,k}$.

The random vector U_I, the families $\{U_{\lambda,k}\}_{0\le k\le d_\lambda}$ for different $\lambda \in \Lambda_{II}$ with $\operatorname{Im}\lambda \ge 0$, and the family $\{W_{\lambda,k}\}_{\lambda\in\Lambda_{III},k\le d_\lambda} \cup \{W\}$ are independent of each other.

The Gaussian random vectors U_I *and* $U_{\lambda,k}$ *are characterized by mean* 0 *and covariances*

$$\mathbb{E}\big(U_I U_I\big) = \Sigma_I,$$
$$\mathbb{E}\big(U_{\lambda,k}U'_{\mu,l}\big) = \frac{1}{(k+l+1)\,k!\,l!} Q_{\lambda,k} N_\lambda^k P_\lambda \Sigma_{II} P'_\mu (N'_\mu)^l Q'_{\mu,l}$$

(which vanishes unless $\mu = \bar{\lambda}$*).* □

Limits at stopping times. The main interest in the functional limit theorem above is that it enables us to study $\mathcal{X}(t)$ at *random* times t. Our main interest is to let t be the n:th splitting time τ_n; as mentioned in the introduction, this yields results for the urn process X_n. As a preparation for this, we will first study another important example where we stop the process when we reach a given total number of balls, or a given number of balls of a given type. Somewhat more generally, let $b \in \mathbb{R}^q$ be a fixed vector and define for $z \geq 0$

$$\tau_b(z) := \min\{t \geq 0 : b \cdot \mathcal{X}(t) \geq z\} \qquad (\text{with } \min\emptyset = +\infty).$$

We assume $b \cdot v_1 > 0$, which means that typically $b \cdot \mathcal{X}(t) \to \infty$, as is shown by the following lemma. A more precise result is given in Lemma 10.1. The proof of this lemma and the following results are given in Section 10.

Lemma 3.9. *Assume* $b \cdot v_1 > 0$. *Conditioned on essential non-extinction, we have a.s.* $b \cdot \mathcal{X}(t) \to \infty$ *as* $t \to \infty$ *and thus* $0 \leq \tau_b(z) < \infty$ *for all* $z \geq 0$. *Moreover,* $\tau_b(z) \to \infty$ *as* $z \to \infty$.

Theorem 3.10. *Assume* (A1)–(A6) *and let* $b \in \mathbb{R}^q$ *with* $b \cdot v_1 > 0$. *Conditioned on essential non-extinction we have as* $z \to \infty$,

$$z^{-1}\mathcal{X}\big(\tau_b(z)\big) \overset{\text{a.s.}}{\to} (b \cdot v_1)^{-1} v_1, \tag{3.4}$$

and, with joint convergence in distribution of all random vectors,

(i) $z^{-1/2} P_I \mathcal{X}\big(\tau_b(z)\big) \overset{\text{d}}{\to} (b \cdot v_1)^{-1/2} V_I$;

(ii) *for every* $\lambda \in \Lambda_{II}$ *and* $k = 0, \dots, d_\lambda$,
$$(z \ln^{2k+1} z)^{-1/2} Q_{\lambda,k} P_\lambda \mathcal{X}\big(\tau_b(z)\big) \overset{\text{d}}{\to} (b \cdot v_1)^{-1/2} V_{\lambda,k};$$

(iii) *for every* $\lambda \in \Lambda_{III}$ *and* $k = 0, \dots, d_\lambda$,
$$(\ln z)^{-k} z^{-\lambda/\lambda_1} Q_{\lambda,k} P_\lambda \mathcal{X}\big(\tau_b(z)\big) \overset{\text{a.s.}}{\to} (b \cdot v_1)^{-\lambda/\lambda_1} \check{W}_{\lambda,k};$$

(iv)
$$z^{-1/2}\left(P_{\lambda_1}\mathcal{X}\big(\tau_b(z)\big) - \frac{z}{b \cdot v_1} v_1 + \sum_{\lambda \neq \lambda_1} \frac{b \cdot P_\lambda \mathcal{X}\big(\tau_b(z)\big)}{b \cdot v_1} v_1\right) \overset{\text{a.s.}}{\to} 0;$$

(v)
$$z^{-1/2}\left(\mathcal{X}\big(\tau_b(z)\big) - \frac{z}{b \cdot v_1} v_1 - \Big(I - \frac{v_1 b'}{b \cdot v_1}\Big) \sum_{\lambda \neq \lambda_1} P_\lambda \mathcal{X}\big(\tau_b(z)\big)\right) \overset{\text{a.s.}}{\to} 0.$$

Here, V_I, $V_{\lambda,k}$ and $\check W_{\lambda,k}$ are vector-valued random variables with V_I and $V_{\lambda,k}$ jointly Gaussian. The vector V_I is real, while the vectors $V_{\lambda,k}$ and $\check W_{\lambda,k}$ are real for real λ but complex otherwise, with $V_{\bar\lambda,k} = \overline{V_{\lambda,k}}$ and $\check W_{\bar\lambda,k} = \overline{\check W_{\lambda,k}}$. Furthermore, a.s., $V_I \in E_I := \bigoplus_{\lambda\in\Lambda_I} E_\lambda$, $V_{\lambda,k} \in E_{\lambda,k}$, and $\check W_{\lambda,k} \in E_{\lambda,k}$.

The random vector V_I, the families $\{V_{\lambda,k}\}_{0\le k\le d_\lambda}$ for different $\lambda \in \Lambda_{II}$ with $\operatorname{Im}\lambda \ge 0$, and the family $\{W_{\lambda,k}\}_{\lambda\in\Lambda_{III},k\le d_\lambda}$ are independent of each other.

The Gaussian vectors V_I and $V_{\lambda,k}$ are characterized by mean 0 and covariances

$$\mathbb{E}\big(V_I V_I'\big) = \Sigma_I, \tag{3.5}$$

$$\mathbb{E}\big(V_{\lambda,k} V_{\mu,l}'\big) = \frac{\lambda_1^{-k-l-1}}{(k+l+1)\,k!\,l!} Q_{\lambda,k} N_\lambda^k P_\lambda \Sigma_{II} P_\mu' (N_\mu')^l Q_{\mu,l}' \tag{3.6}$$

(which vanishes unless $\mu = \bar\lambda$).

The same results hold for $X_{\tau_b(z)}$ in the urn process.

We specialize to some important cases.

Corollary 3.11. *Assume* (A1)–(A6) *and let $b \in \mathbb{R}^q$ with $b\cdot v_1 > 0$. Suppose further $\operatorname{Re}\lambda_2 < \frac12\lambda_1$. Conditioned on essential non-extinction we have as $z \to \infty$,*

$$z^{-1/2}\Big(\mathcal{X}\big(\tau_b(z)\big) - \frac{z}{b\cdot v_1} v_1\Big) \overset{\mathrm{d}}{\to} N(0,\Sigma_b),$$

where the covariance matrix Σ_b is given by

$$\begin{aligned}\Sigma_b &= (b\cdot v_1)^{-1}\Big(I - \frac{v_1 b'}{b\cdot v_1}\Big)\Sigma_I\Big(I - \frac{b v_1'}{b\cdot v_1}\Big)\\ &= (b\cdot v_1)^{-1}\int_0^\infty \Big(I - \frac{v_1 b'}{b\cdot v_1}\Big) e^{sA} B e^{sA'}\Big(I - \frac{b v_1'}{b\cdot v_1}\Big) e^{-\lambda_1 s}\,ds.\end{aligned} \tag{3.7}$$

The same result holds for $X_{\tau_b(z)}$ in the urn process.

Corollary 3.12. *Assume* (A1)–(A6) *and let $b \in \mathbb{R}^q$ with $b\cdot v_1 > 0$. Suppose further $\operatorname{Re}\lambda_2 = \frac12\lambda_1$, and let $d := \max\{d_\lambda : \operatorname{Re}\lambda = \frac12\lambda_1\}$. Conditioned on essential non-extinction we have as $z \to \infty$,*

$$\big(z\ln^{2d+1} z\big)^{-1/2}\Big(\mathcal{X}\big(\tau_b(z)\big) - \frac{z}{b\cdot v_1} v_1\Big) \overset{\mathrm{d}}{\to} N(0,\Sigma_b),$$

where the covariance matrix Σ_b is given by

$$\Sigma_b = \frac{\lambda_1^{-2d-1}}{b\cdot v_1}\Big(I - \frac{v_1 b'}{b\cdot v_1}\Big)\Sigma_{II,d}\Big(I - \frac{b v_1'}{b\cdot v_1}\Big). \tag{3.8}$$

The same result holds for $X_{\tau_b(z)}$ in the urn process.

Corollary 3.13. *Assume* (A1)–(A6) *and let $b \in \mathbb{R}^q$ with $b\cdot v_1 > 0$. Suppose further $\operatorname{Re}\lambda_2 > \frac12\lambda_1$, and let $d := \max\{d_\lambda : \operatorname{Re}\lambda = \operatorname{Re}\lambda_2\}$. Conditioned on essential non-extinction, the family of random variables*

$$Y_b(z) := (\ln z)^{-d} z^{-\operatorname{Re}\lambda_2/\lambda_1}\Big(\mathcal{X}\big(\tau_b(z)\big) - \frac{z}{b\cdot v_1} v_1\Big),$$

for $z \geq 2$, say, is tight. More precisely, there exist complex random vectors $W_{b,\lambda}$, $\lambda \in \Lambda'_{III} := \{\lambda : \operatorname{Re}\lambda = \operatorname{Re}\lambda_2, \operatorname{Im}\lambda \geq 0$ and $d_\lambda = d\}$, such that, as $z \to \infty$,

$$Y_b(z) - \operatorname{Re} \sum_{\lambda \in \Lambda'_{III}} e^{-\mathrm{i}(\operatorname{Im}\lambda/\lambda_1)\ln z} W_{b,\lambda} \overset{\text{a.s.}}{\to} 0. \tag{3.9}$$

In particular, if λ_2 is real and $\Lambda'_{III} = \{\lambda_2\}$, then as $z \to \infty$, $Y_b(z) - W_b \overset{\text{a.s.}}{\to} 0$ for some random vector W_b, and thus

$$Y_b(z) \overset{\text{d}}{\to} W_b. \tag{3.10}$$

The same results hold for $X_{\tau_b(z)}$ in the urn process.

Remark 3.14. Unless all ξ_i a.s. belong to some fixed subspace of $\mathbb{R}^q$, the matrix B has full rank q, and thus the covariance matrix in (3.7) has rank $q-1$; hence the limit distribution is concentrated on the hyperplane $\{v : b \cdot v = 0\}$ but not on any smaller subspace.

In contrast, the limits in Corollaries 3.12 and 3.13 are typically concentrated on a subspace of $\mathbb{R}^q$ of low dimension (commonly 1 or 2).

Remark 3.15. Suppose that $\operatorname{Re}\lambda_2 > \frac{1}{2}\lambda_1$ with $\operatorname{Im}\lambda_2 > 0$, and that, for simplicity, $\Lambda'_{III} = \{\lambda_2\}$. In that case, the sum in (3.9) contains only one term, and we see that $Y_b(z)$ converges in distribution when $z \to \infty$ along a subsequence where the fractional part of $(\operatorname{Im}\lambda_2/2\pi\lambda_1)\ln z$ converges. The subsequence limits are of the form $\operatorname{Re}\bigl(e^{\mathrm{i}\alpha} W_b\bigr)$ for a complex random vector W_b and real α. Hence, a limit distribution as in (3.10) exists only if $\omega W_b \overset{\text{d}}{=} W_b$ whenever $|\omega| = 1$, i.e. if W_b is symmetric complex. This seems highly unlikely, and we conjecture that it never happens, so we do not expect that Y_b converges in distribution in this case (or, more generally, when Λ'_{III} contains a non-real λ). Unfortunately, we have not been able to show this conjecture in general, but in some particular cases [25], [16], non-convergence of $Y_b(z)$ has been shown by computations of the first moments, showing that the moments oscillate.

Similarly, we conjecture that the (subsequence) limits in Corollary 3.13 never are normal; again we cannot prove this, but it can be verified in some particular cases by computations of moments [14], [49], [25], [16].

Limits for urns. In some urn processes we add a fixed number of balls each time (with varying types), say m; assuming as we may that we start with a non-random number of balls l, the total number of balls at time n then is deterministic $mn + l$ and τ_n equals $\tau_b(mn + l)$ with $b = (1, 1, \dots, 1)'$ (or $\tau_b(n + l/m)$ with $b = m^{-1}(1, 1, \dots, 1)'$). In this case, Theorem 3.10 and its corollaries thus yield results for $\mathcal{X}(\tau_n)$ and thus for X_n.

In general, we can change the setup by adding a dummy type $q + 1$ such that $a_{q+1} = 0$ (the dummy balls never split) and a dummy ball is added whenever a ball splits (see Section 10 for details). The dummy balls then count the number of splits, and Theorem 3.10 yields a description of the asymptotics of X_n. In order to give explicit expressions for the asymptotic variances, we define, for $s \geq 0$,

$$\phi(s,A) := \sum_{n=1}^{\infty} \frac{s^n}{n!} A^{n-1} = \int_0^s e^{tA} dt, \tag{3.11}$$

$$\psi(s,A) := e^{sA} - \lambda_1 v_1 a' \phi(s,A). \tag{3.12}$$

The first result below is in [9, Section V.9.3] (under slightly different hypotheses), and is included for completeness.

Theorem 3.16. *Assume* (A1)–(A6). *Conditioned on essential non-extinction, $n^{-1} X_n \overset{\text{a.s.}}{\to} \lambda_1 v_1$ as $n \to \infty$.*

Theorem 3.17. *Assume* (A1)–(A6). *Suppose further* $\operatorname{Re}\lambda_2 < \frac{1}{2}\lambda_1$. *Conditioned on essential non-extinction we have as $n \to \infty$,*

$$n^{-1/2}\big(X_n - n\lambda_1 v_1\big) \overset{\text{d}}{\to} N(0,\Sigma),$$

where the covariance matrix Σ is given by

$$\Sigma = \int_0^\infty \psi(s,A) B \psi(s,A)' e^{-\lambda_1 s} \lambda_1 \, ds - \lambda_1^2 v_1 v_1'. \tag{3.13}$$

Theorem 3.18. *Assume* (A1)–(A6). *Suppose further* $\operatorname{Re}\lambda_2 = \frac{1}{2}\lambda_1$, *and let $d := \max\{d_\lambda : \operatorname{Re}\lambda = \frac{1}{2}\lambda_1\}$. Conditioned on essential non-extinction we have as $z \to \infty$,*

$$\big(n \ln^{2d+1} n\big)^{-1/2}\big(X_n - n\lambda_1 v_1\big) \overset{\text{d}}{\to} N(0,\Sigma),$$

where the covariance matrix Σ is given by

$$\Sigma = \lambda_1^{-2d}(I-T)\Sigma_{II,d}(I-T'), \tag{3.14}$$

with $T := \sum_{\lambda \in \Lambda_{II}} \lambda^{-1}\lambda_1 v_1 a' P_\lambda$. If $a \in \operatorname{Im}(A')$ and $a = A'\hat{a}$, T can be replaced by $T_1 := \lambda_1 v_1 \hat{a}'$.

Theorem 3.19. *Assume* (A1)–(A6). *Suppose further* $\operatorname{Re}\lambda_2 > \frac{1}{2}\lambda_1$, *and let $d := \max\{d_\lambda : \operatorname{Re}\lambda = \operatorname{Re}\lambda_2\}$. Conditioned on essential non-extinction, the family of random variables*

$$Y(n) := \big(n \ln^{2d} n\big)^{-1/2}\big(X_n - n\lambda_1 v_1\big), \qquad n \geq 2,$$

is tight, and we have the same type of asymptotic behaviour as described in Corollary 3.13 and Remark 3.15. In particular, if λ_2 is real and $\operatorname{Re}\lambda_3 < \lambda_2$, *or more generally when $\Lambda'_{III} = \{\frac{1}{2}\lambda_1\}$, then $Y(n) \overset{\text{d}}{\to} W_b$ for some random vector W_b, but we conjecture that this fails otherwise.*

Remark 3.20. In Theorem 3.18, the limit is typically concentrated on a subspace L of low dimension (as in Remark 3.14). For some vectors $\eta \in \mathbb{R}^q$ with $\eta \perp L$ (and thus $\eta\Sigma\eta' = 0$) it is possible to obtain a non-degenerate limit in distribution of $\eta \cdot X_n$ with a different normalization from Theorem 3.10(i) or (ii); we leave the details to the reader. The same applies to Theorem 3.19 (and subsequence limits). Cf. [8], [9].

The drawn balls. Let N_{ni} be the number of drawn balls of type i in the first n draws, and let $N_n := (N_{n1}, \dots, N_{nq})$. (In the branching process, we would similarly study $\mathcal{N}_i(t)$, the number of deaths of balls of type i up to time t. We leave this case to the reader.)

If each ξ_i is deterministic and, for notational simplicity, each $a_i = 1$, then $X_n = X_0 + AN_n$, so if further A is invertible, asymptotics for $N_n = A^{-1}(X_n - X_0)$ follow from Theorems 3.16–3.19. In general, we can argue with dummy balls again, this time using q dummy types $q+1, \dots, 2q$ and adding a dummy ball of type $q+i$ each time a ball of type i is drawn. This leads to the following theorem. (Explicit expressions when some $a_i \neq 1$ can be found by this method too, but are left to the reader.)

Theorem 3.21. *Assume* (A1)–(A6). *Conditioned on essential non-extinction we have, as* $n \to \infty$, $n^{-1} N_{ni} \xrightarrow{\text{a.s.}} \nu_i := a_i v_{1i}$ *and, furthermore,*

(i) *if* $\operatorname{Re}\lambda_2 < \lambda_1/2$, *then*

$$n^{-1/2}\big(X_n - n\lambda_1 v_1,\, N_n - n\nu\big) \xrightarrow{\text{d}} (V, \widehat{V}),$$

where $(V, \widehat{V})$ *is vector-valued Gaussian random variable with mean* 0*; if for simplicity each* $a_i = 1$, *and* D_v *is the diagonal matrix with entries* $D_{ii} = v_{1i}$,

$$\mathbb{E}\,\widehat{V}\widehat{V}' = \int_0^\infty (I - v_1 a')\Big(\phi(s,A)B\phi(s,A)' + e^{sA}D_v + D_v e^{sA'} - D_v\Big)(I - a v_1')e^{-\lambda_1 s}\lambda_1\,ds. \tag{3.15}$$

$$\mathbb{E}\,V\widehat{V}' = \int_0^\infty \psi(s,A)\Big(B\phi(s,A)' + AD_v\Big)(I - a v_1')e^{-\lambda_1 s}\lambda_1\,ds. \tag{3.16}$$

(ii) *if* $\operatorname{Re}\lambda_2 = \lambda_1/2$ *and* $d := \max\{d_\lambda : \operatorname{Re}\lambda = \frac{1}{2}\lambda_1\}$, *then*

$$\big(n\ln^{2d+1} n\big)^{-1/2}\big(X_n - n\lambda_1 v_1,\, N_n - n\nu\big) \xrightarrow{\text{d}} (V, \widehat{V}),$$

where $(V, \widehat{V})$ *is a vector-valued Gaussian random variable with mean* 0. *If, for simplicity, each* $a_i = 1$, *then* $V = A\widehat{V}$ *and*

$$\mathbb{E}\big(\widehat{V}\widehat{V}'\big) = \lambda_1^{-2d}\widehat{T}\Sigma_{II,d}\widehat{T}',$$

with $\widehat{T} := (I - v_1 a')\sum_{\lambda\in\Lambda_{II}}\lambda^{-1}P_\lambda$.

In the case of random ξ_i, the dummy ball method also shows asymptotic normality when $\lambda_2 \le \lambda_1/2$ of the number of draws of a ball of type i leading to a specific set of balls being added. We leave the details to the reader.

Functional limit theorems for stopped processes and urns.

Theorem 3.22. *Assume* (A1)–(A6) *and let* $b \in \mathbb{R}^q$ *with* $b \cdot v_1 > 0$. *Conditioned on essential non-extinction we have as* $z \to \infty$, *with joint convergence,*

(i) $z^{-1/2}P_I\mathcal{X}\big(\tau_b(xz)\big) \xrightarrow{\text{d}} (b\cdot v_1)^{-1/2}V_I(x)$ *in* $D[0,\infty)$*;*

(ii) *for every* $\lambda \in \Lambda_{II}$ *and* $k = 0, \dots, d_\lambda$, *in* $D[0,\infty)$,

$$(\ln z)^{-k-1/2}(z^x)^{-1/2-\mathrm{i}\operatorname{Im}\lambda/\lambda_1} Q_{\lambda,k} P_\lambda \mathcal{X}\bigl(\tau_b(z^x)\bigr) \overset{\mathrm{d}}{\to} (b\cdot v_1)^{-1/2}\lambda_1^{-(k+1/2)} U_{\lambda,k}(x).$$

Here V_I *and* $U_{\lambda,k}$ *are continuous Gaussian vector-valued stochastic processes, defined on* $[0,\infty)$. $U_{\lambda,k}(x)$ *is as in Theorem 3.1,* $V_I(0) = 0$, $\mathbb{E}\, V_I(x) = 0$ *and*

$$\mathbb{E}\, V_I(x) V_I(y)' = x \Sigma_I \Bigl(\frac{y}{x}\Bigr)^{\lambda_1^{-1} A'}, \qquad 0 < x \le y. \tag{3.17}$$

The process V_I *and the families* $\{U_{\lambda,k}\}_{0\le k\le d_\lambda}$ *for different* $\lambda \in \Lambda_{II}$ *with* $\operatorname{Im}\lambda \ge 0$ *are independent.*

The same results hold for $X_{\tau_b(xz)}$ *and* $X_{\tau_b(z^x)}$ *in the urn process.*

Corollary 3.23. *Assume* (A1)–(A6) *and let* $b \in \mathbb{R}^q$ *with* $b \cdot v_1 > 0$. *Let* $z \to \infty$ *and condition on essential non-extinction.*

(i) *If* $\operatorname{Re}\lambda_2 < \frac12\lambda_1$, *then, in* $D[0,\infty)$,

$$z^{-1/2}\Bigl(\mathcal{X}\bigl(\tau_b(xz)\bigr) - \frac{xz}{b\cdot v_1} v_1\Bigr) \overset{\mathrm{d}}{\to} V_b(x) := (b\cdot v_1)^{-1/2}\Bigl(I - \frac{v_1 b'}{b\cdot v_1}\Bigr) V_I(x).$$

(ii) *If* $\lambda_2 = \frac12\lambda_1$ *and* $\Lambda_{II} = \{\lambda_2\}$, *then, with* $d := d_{\lambda_2}$, *in* $D[0,\infty)$,

$$(\ln z)^{-d-1/2} z^{-x/2}\Bigl(\mathcal{X}\bigl(\tau_b(z^x)\bigr) - \frac{z^x}{b\cdot v_1} v_1\Bigr) \overset{\mathrm{d}}{\to}$$
$$V_b(x) := \lambda_1^{-d-1/2}(b\cdot v_1)^{-1/2}\Bigl(I - \frac{v_1 b'}{b\cdot v_1}\Bigr) U_{\lambda_2,d}(x).$$

The limit processes are Gaussian, vanish at 0, *and have means* 0 *and covariances given by* (3.17) *for* (i) *and, using* (3.3), *for* (ii)

$$\mathbb{E}\, U_{\lambda_2,d}(x) U_{\lambda_2,d}(y)' = c(d,d,x,y)(2d+1) d!^2 \Sigma_{II,d}, \qquad 0 \le x \le y. \tag{3.18}$$

The same results holds for the urn process.

With dummy balls as above, Corollary 3.23 leads to the corresponding result for urns.

Theorem 3.24. *Assume* (A1)–(A6). *Let* $n \to \infty$ *and condition on essential non-extinction.*

(i) *If* $\operatorname{Re}\lambda_2 < \frac12\lambda_1$, *then, in* $D[0,\infty)$,

$$n^{-1/2}\bigl(X_{\lfloor xn\rfloor} - xn\lambda_1 v_1\bigr) \overset{\mathrm{d}}{\to} V(x),$$

where $V(x)$ *is a continuous Gaussian vector-valued process with* $V(0) = 0$, *mean* $\mathbb{E}\, V(x) = 0$ *and, for* $0 < x \le y$,

$$\mathbb{E}\, V(x) V'(y) = \int_{-\lambda_1^{-1}\ln x}^{\infty} \psi(s + \lambda_1^{-1}\ln x, A) B \psi(s + \lambda_1^{-1}\ln y, A)' e^{-\lambda_1 s}\lambda_1 \, ds - x\lambda_1^2 v_1 v_1'. \tag{3.19}$$

(ii) *If* $\lambda_2 = \frac{1}{2}\lambda_1$ *and* $\Lambda_{II} = \{\lambda_2\}$*, then, with* $d := d_{\lambda_2}$*, in* $D[0,\infty)$*,*

$$(\ln n)^{-d-1/2} n^{-x/2}\big(X_{\lfloor n^x \rfloor} - n^x \lambda_1 v_1\big) \xrightarrow{\mathrm{d}} V(x),$$

where $V(x)$ *is a continuous Gaussian vector-valued process with* $V(0) = 0$*, mean* $\mathbb{E}\, V(x) = 0$ *and, for* $0 < x \le y$*,*

$$\mathbb{E}\, V(x)V(y)' = \tilde{c}(d,x,y)\lambda_1^{-2d}(I - 2v_1 a' P_{\lambda_2})\Sigma_{II,d}(I - 2P'_{\lambda_2} a v'_1), \quad (3.20)$$

where

$$\tilde{c}(d,x,y) := (2d+1)d!^2 c(d,d,x,y) = \sum_{j=0}^{d} \frac{2d+1}{2d+1-j}\binom{d}{j}(y-x)^j x^{2d+1-j}.$$

Remark 3.25. If $\operatorname{Re}\lambda_2 = \frac{1}{2}\lambda_1$ and $\operatorname{Im}\lambda_2 \neq 0$ we obtain a more complicated behaviour. If, for simplicity, λ_2 and $\overline{\lambda_2}$ are the only eigenvalues with real part $\frac{1}{2}\lambda_1$, it follows from Theorem 3.22 that $\big(X_{\lfloor n^x \rfloor} - n^x \lambda_1 v_1\big)$ will, asymptotically, oscillate deterministically as a sine function with frequency $(\operatorname{Im}\lambda_2/2\pi\lambda_1)\ln n$, but with amplitude and phase drifting stochastically at a slower rate.

4. Some remarks, extensions and open problems

Remark 4.1. We assume throughout that the set of types is finite. However, in Examples 7.6 and 7.7 we consider two applications where the natural urn models have an *infinite* number of types (in this case $\mathbb{N}$); luckily it is possible in those applications to consider only a finite number of types at a time.

Another example where the results extend to an infinite space of types (in that case a compact group, for example $\mathbb{T}$) is described in Example 7.10.

These examples suggest the possibility of (and desire for) an extension of the results in this paper to infinite sets of types (with suitable assumptions). Our matrix A would then be replaced by an operator acting in a suitable space, such as $\ell^1(\mathbb{N})$ or $L^2(\mathbb{T})$. It is far from clear how such an extension should be formulated, and we have not pursued this.

Remark 4.2. In several applications, the assumptions (1.1), (1.2) are too restrictive; we want to allow the possibility of removing other balls than the drawn one. Several authors, following Bagchi and Pal [10], have studied the so-called *tenable* urn models where ξ_{ii} may be an arbitrary negative integer $-d_i$, but it is assumed that $d_i | \xi_{ji}$ for all j and $d_i | X_{0i}$; hence X_{ni} is always a multiple of d_i and we can never be required to remove balls that do not exist. (We still assume $\xi_{ji} \ge 0$ when $j \neq i$. We let $d_i = 1$ if $\xi_{ii} \ge 0$.)

Note that the corresponding continuous time process $\mathcal{X}(t)$ is well-defined, but (if some $d_i \ge 2$) it is not a branching process because the balls do not evolve independently.

Tenable urns can, nevertheless, easily be reduced to the Athreya–Karlin setting, and thus studied by the results above, by replacing the balls with "superballs"; each superball of type i being equivalent to d_i ordinary balls and having activity $d_i a_i$. If D denotes the diagonal matrix with $D_{ii} = d_i$, this means that we consider the process $\widetilde{X}_n := D^{-1}X_n$, which is of the type treated above with $\widetilde{\xi}_i = D^{-1}\xi_i$ and $\widetilde{A} = D^{-1}AD$. Note that $\widetilde{A}$ and A have the same eigenvalues. If (A2)–(A6) hold, we can thus apply Theorem 3.1 and its consequences in Section 3 to the superball process. Returning to the original process (by multiplying with D), it can easily be verified that all results in Section 3 (including the variance formulas) hold for tenable urns too. See Example 7.5 for an example.

As mentioned in the introduction, the results can be extended even further. Let ξ_{ij} be arbitrary integers (or even real numbers), but assume that they are such that the urn process never can require the removal of balls that do not exist. (An example, from [39], is given in Example 7.8.) The corresponding continuous time process then is well-defined; it is a generalized branching process where the death of one ball may force the removal of others. In this case, we cannot use the Perron–Frobenius theory, so we add the assumption that A has a real eigenvalue $\lambda_1 > 0$ with $\operatorname{Re}\lambda < \lambda_1$ for every other eigenvalue λ, and that there exist corresponding left and right eigenvectors u_1 and v_1 such that $v_{1i} > 0$ for every i and $u_{1i} > 0$ if i is dominating while $u_{1i} = 0$ otherwise; we also assume (A2)–(A6). Finally, we assume that Lemma 9.7(iii) holds (we have not been able to prove this in the present generality), for example because $\mathbb{P}(W = 0) = 0$. It may then be verified that the proofs in Sections 9 and 10 hold without modification; hence all results in Section 3 hold for such urn models and generalized branching processes too. (We conjecture that the results hold also if it is possible that the process stops by requiring some X_{ni} to become negative, provided that we condition on this not happening, but we have not pursued this.)

Remark 4.3. Assume for simplicity that essential extinction is impossible, so $W > 0$. By Theorem 3.1 or Corollary 3.8, the different sets of projections of $\mathcal{X}(t)$ in parts (i), (ii) and (iii) of these results, divided by the normalizing factor $W^{1/2}$, form three asymptotically independent families. This may seem surprising at first sight, but is explained by the three families being essentially determined by what happens in the end, the middle, and the beginning of the process, respectively. More formally, let $\varepsilon(t) \to 0$ and $\omega(t) \to \infty$ as $t \to \infty$. By Theorem 3.1 (and its proof), dividing everything by $W^{1/2}$ and ignoring terms that are asymptotically negligible, the projections in (i) depend only on the random splits after time $t-\omega(t)$, the projections in (ii) depend only on the random splits in $[\varepsilon(t)t, (1-\varepsilon(t))t]$, and the projections in (iii) depend only on the random splits before $\omega(t)$.

For the urn process in Theorems 3.17–3.19 we find, by the exponential relation between n and t, that the variables asymptotically depend only on the draws after εn, the draws in $[n^\varepsilon, n^{1-\varepsilon}]$ and the draws before ω, respectively. The same holds if we consider the different cases together as in Remark 3.20.

Remark 4.4. It would be interesting to extend the results to cases where (A4) or (A6) does not hold. A typical case is when A is triangular: if, for example, $q = 2$ and $A = \left(\begin{smallmatrix} \alpha & 0 \\ \beta & \delta \end{smallmatrix}\right)$ with $\beta > 0$, the conditions hold if $\alpha > \delta$ but not otherwise. It might be possible to handle this case by combining the methods here with the ones in [34], where a detailed study is made in the related case of multitype Galton–Watson processes in discrete time. Some new phenomena will arise, however, see [34], [20] and [35].

Remark 4.5. Mahmoud [40] has initiated the study of urn models where several, say 2, balls are drawn at the same time, and balls are added depending on the drawn combination of types. It may be possible to study such models too by the methods of this paper, first considering the corresponding continuous time model, but we have not pursued this. (This case is substantially more complicated than the standard case treated here; for example, the continuous time model will explode in finite time.)

Remark 4.6. Asmussen [3] has proved laws of iterated logarithm for $\mathcal{X}(t)$ in the irreducible case. (The results in the cases $\operatorname{Re}\lambda_2 < \frac{1}{2}\operatorname{Re}\lambda_1$ and $\operatorname{Re}\lambda_2 = \frac{1}{2}\operatorname{Re}\lambda_1$ are different.) By the Athreya–Karlin embedding, this yields laws of iterated logarithm for the urn process X_n complementing the results above; we leave the details to the reader. Such laws for the urn process have been proved in a special case by Bai, Hu and Zhang [12].

Remark 4.7. Our methods give no information on the rate of convergence. Using other methods, Hwang [23] has found the rate of convergence to the limiting normal distribution for a specific variable in Example 7.9 below: the rate is $n^{-\gamma}$ where $\gamma = \min\bigl(\frac{1}{2}, 3(\frac{1}{2} - \operatorname{Re}\lambda_2/\lambda_1)\bigr)$. It is tempting to guess that this might hold rather generally.

5. Variance calculations

In several of the theorems in Section 3, the variances and covariances of the limits are given as integrals of matrix functions. In any specific application, these integrals can be evaluated by first transforming the matrix A to Jordan normal form. (A computer algebra package is helpful, and can do the integration directly if q is not too large.) We will here give some simplifications in important special cases. We concentrate on the main cases; the reader may add further similar results. See the examples in Section 7 and for various applications of these results. See also [29] for more complicated applications.

First we consider the case when the replacement vectors ξ_i are deterministic. Let D be the diagonal $q \times q$ matrix with entries

$$D_{ii} := \begin{cases} v_{1i}/a_i, & a_i \neq 0, \\ 0, & a_i = 0. \end{cases} \tag{5.1}$$

Lemma 5.1. (i) *If each ξ_i is deterministic, then $B = ADA'$.*

(ii) *If furthermore* $\operatorname{Re}\lambda_2 < \frac{1}{2}\lambda_1$, *then the covariance matrix* Σ *in Theorem 3.17 is given by*

$$\Sigma = \int_0^\infty (A - \lambda_1 v_1 a')e^{sA} D e^{sA'}(A - \lambda_1 v_1 a')' e^{-\lambda_1 s}\lambda_1 \, ds. \tag{5.2}$$

Equivalently, if $g_i(s) := (A - \lambda_1 v_1 a')e^{sA}\delta_i$, *where* $(\delta_i)_j = \delta_{ij}$, *then*

$$\Sigma = \int_0^\infty \sum_{a_i \neq 0} a_i^{-1} v_{1i} g_i(s) g_i(s)' e^{-\lambda_1 s}\lambda_1 \, ds. \tag{5.3}$$

If further $u_1 = a$, *then* $g_i(s) = Ae^{sA}(I - v_1 u_1')\delta_i = Ae^{sA}(\delta_i - u_{1i}v_1)$.

Proof. We now have, by (2.13) and (2.1), when $a_i \neq 0$,

$$(B_i)_{jk} = \mathbb{E}\,\xi_{ij}\,\mathbb{E}\,\xi_{ik} = a_i^{-2} A_{ji} A_{ki},$$

and then by (2.14)

$$B_{jk} = \sum_{i=1}^q a_i v_{1i} (B_i)_{jk} = \sum_{a_i \neq 0} a_i^{-1} v_{1i} A_{ji} A_{ki} = (ADA')_{jk},$$

which proves (i).

For (ii), we have by (3.11) and (3.12) $\phi(s, A)A = e^{sA} - 1$ and

$$\psi(s, A)A = e^{sA}A - \lambda_1 v_1 a'(e^{sA} - 1) = (A - \lambda_1 v_1 a')e^{sA} + \lambda_1 v_1 a'.$$

Hence (i) now yields

$$\begin{aligned}\psi(s, A)B\psi(s, A)' = (A - \lambda_1 v_1 a')e^{sA} D e^{sA'}(A - \lambda_1 v_1 a')' \\ + \lambda_1 v_1 a' D e^{sA'}(A - \lambda_1 v_1 a')' + \lambda_1 (A - \lambda_1 v_1 a')e^{sA} D a v_1' \\ + \lambda_1^2 v_1 a' D a v_1'.\end{aligned}$$

It follows from (5.1) that $(Da)_i = v_{1i}$ when $a_i \neq 0$, and thus $a'Da = a'v_1 = 1$ and $ADa = Av_1 = \lambda_1 v_1$. It follows that $(A - \lambda_1 v_1 a')e^{sA}(Da - v_1) = 0$, and thus $(A - \lambda_1 v_1 a')e^{sA} Da = (A - \lambda_1 v_1 a')e^{sA} v_1 = 0$. Hence, the second and third terms in the sum vanish. The fourth equals $\lambda_1^2 v_1 v_1'$, and thus (5.2) follows from (3.13). Since $D = \sum_i D_{ii}\delta_i\delta_i'$, (5.3) follows from (5.2).

In the special case $u_1 = a$, we have $(A - \lambda_1 v_1 a')e^{sA} = Ae^{sA}(I - v_1 u_1')$, and the alternative formulas for g_i follow. □

Remark 5.2. In general, the argument above shows that if B_i° is the covariance matrix $\mathbb{E}(\xi_i - \mathbb{E}\,\xi_i)(\xi_i - \mathbb{E}\,\xi_i)'$ and $B^\circ := \sum_{i=1}^q a_i v_{1i} B_i^\circ$, then

$$\begin{aligned}\Sigma = \int_0^\infty \psi(s, A) B^\circ \psi(s, A)' e^{-\lambda_1 s}\lambda_1 \, ds \\ + \int_0^\infty (A - \lambda_1 v_1 a')e^{sA} D e^{sA'}(A - \lambda_1 v_1 a')' e^{-\lambda_1 s}\lambda_1 \, ds,\end{aligned}$$

which separates the contributions to the asymptotic variance coming from the randomness in the ξ_i and the randomness in the draws.

Another simplifying case is when A is diagonalizable. In that case, there exist dual bases $\{u_i\}_{i=1}^q$ and $\{v_i\}_{i=1}^q$ of left and right eigenvectors of A, i.e. vectors such that $u_iA = \lambda_i u_i$, $Av_i = \lambda_i v_i$ and $u_i \cdot v_j = \delta_{ij}$ (where the λ_i, $i = 1, \dots, q$ do not have to be distinct; we assume that the bases are ordered such that $\lambda_1 \geq \operatorname{Re}\lambda_2 \geq \dots$ as elsewhere).

Lemma 5.3. (i) *If A is diagonalizable and $\{u_i\}_{i=1}^q$ and $\{v_i\}_{i=1}^q$ are dual bases of eigenvectors, then, with $L_I := \{i : \lambda_i \in \Lambda_I\}$ and $L_{II} := \{i : \lambda_i \in \Lambda_{II}\}$,*

$$\Sigma_I = \sum_{j,k\in L_I} \frac{u_j' B u_k}{\lambda_1 - \lambda_j - \lambda_k} v_j v_k' \qquad \text{and} \qquad \Sigma_{II} = \sum_{j\in L_{II}} (u_j' B \overline{u_j}) v_j v_j^*.$$

(ii) *If further each ξ_i is deterministic, then*

$$\Sigma_I = \sum_{j,k\in L_I} \frac{\lambda_j\lambda_k u_j' D u_k}{\lambda_1 - \lambda_j - \lambda_k} v_j v_k' \qquad \text{and} \qquad \Sigma_{II} = \sum_{j\in L_{II}} |\lambda_j|^2 (u_j' D \overline{u_j}) v_j v_j^*.$$

(iii) *If further $\operatorname{Re}\lambda_2 < \frac{1}{2}\lambda_1$, then the covariance matrix Σ in Theorem 3.17 is given by, with $w_j := \lambda_j v_j - \lambda_1 (a\cdot v_j) v_1$,*

$$\Sigma = \sum_{j,k=2}^{q} \frac{\lambda_1\, u_j' D u_k}{\lambda_1 - \lambda_j - \lambda_k} w_j w_k'.$$

(iv) *If the assumptions in* (i) *and* (ii) *hold and $\operatorname{Re}\lambda_2 = \frac{1}{2}\lambda_1$, then the covariance matrix Σ in Theorem 3.18 is given by, with w_j as in* (iii),

$$\Sigma = \sum_{j\in\Lambda_{II}} (\, u_j' D u_j) w_j w_j^*.$$

Proof. We have $P_I = \sum_{j\in L_I} v_j u_j'$ and $P_I e^{sA} = \sum_{j\in L_I} e^{s\lambda_j} v_j u_j'$. Hence the first equality in (i) follows from the definition (2.15). The second follows similarly from (2.16).

For (ii) we use (i) and $B = ADA'$ from Lemma 5.1, recalling that $u_j' A = \lambda_j u_j'$.

In case (iii), when $L_I = \{2, \dots, q\}$, the result follows from (5.2) and

$$(A - \lambda_1 v_1 a')e^{sA} = \sum_{j\geq 1} (A - \lambda_1 v_1 a') e^{sA} v_j u_j' = \sum_{j\geq 2} w_j e^{s\lambda_j} u_j'.$$

Finally, since A is diagonalizable, $d = 0$ and (iv) follows from (3.14) and (ii). ☐

A third common simplifying case is when a fixed number of balls is added each time, i.e. each $\sum_j \xi_{ij} = m$ is deterministic and independent of i, and further each $a_i = 1$. We state the result somewhat more generally.

Lemma 5.4. *Suppose that $a \cdot \mathbb{E}\,\xi_i = m$ for some $m > 0$ and every i. (For example, this holds if exactly m balls are added each time, and each $a_i = 1$.) Then $\lambda_1 = m$ and $u_1 = a$. The covariance matrix* (3.13) *in*

Theorem 3.17 equals $m\Sigma_I$. *The covariance matrix* (3.14) *in Theorem 3.18 equals* $\lambda_1^{-2d}\Sigma_{II,d}$.

Proof. We have, by (2.1) and assumption,

$$(a'A)_k = \sum_{j=1}^{q} a_j A_{jk} = \sum_{j=1}^{q} a_j a_k \operatorname{\mathbb{E}} \xi_{kj} = a_k a \cdot \operatorname{\mathbb{E}} \xi_k = m a_k$$

so $a'A = ma'$ and a is a non-negative left eigenvector of A. This implies that $m = \lambda_1$ and that a is a multiple of u_1; by our normalizations (2.2) and (2.3), $a = u_1$.

It follows that $a'P_{\lambda_1} = a'$ and $a'P_\lambda = a'P_{\lambda_1}P_\lambda = 0$ when $\lambda \neq \lambda_1$; thus $T = 0$ in Theorem 3.18, which shows the claim about (3.14).

For (3.13), we may use algebraic manipulations as in the lemmas above, but it seems easier to proceed as follows. Conditioning on X_0, we may assume that X_0 is fixed. Then $a \cdot X_n = mn + l$, where $l := a \cdot X_0$. Thus we can obtain X_n by stopping at $\tau_b(n + l/m)$, where $b = m^{-1}a$. We thus obtain the conclusion of Theorem 3.17 directly from Corollary 3.11, with Σ given by (3.7). Furthermore, since $b'P_\lambda = m^{-1}a'P_\lambda = 0$ for every $\lambda \neq \lambda_1$, we have $b'P_I = 0$ and $P_I'b = 0$. Moreover, $b \cdot v_1 = m^{-1}a \cdot v_1 = m^{-1}$. Since $\Sigma_I = P_I\Sigma_I P_I'$ by (2.15), $b'\Sigma_I = 0$ and $\Sigma_I b = 0$, and the middle expression in (3.7) equals $m\Sigma_I$. ∎

If further $m = 1$, we have in addition the following.

Lemma 5.5. *If each* $|\xi_i| = 1$, *i.e. exactly one ball is added each time, and each* $a_i = 1$, *then* $B = D$.

Proof. Since only one ξ_{ij} is non-zero at a time, B_i in (2.13) is diagonal, with $(B_i)_{jj} = \mathbb{P}(\xi_{ij} = 1) = \operatorname{\mathbb{E}} \xi_{ij}$. Thus B is diagonal with, see (2.18),

$$B_{jj} = \sum_{i=1}^{q} v_{1i} a_i \operatorname{\mathbb{E}} \xi_{ij} = (\lambda_1 v_1)_j = v_{1j}. \qquad \square$$

We also give a calculation of the variances and covariances in Theorem 3.21.

Lemma 5.6. *Suppose that* $a_i = 1$ *and* $\operatorname{\mathbb{E}} \sum_j \xi_{ij} = m$ *for every* i, *and that* A *is diagonalizable with dual bases of eigenvectors* $\{u_i\}_{i=1}^q$ *and* $\{v_i\}_{i=1}^q$. *Then*

$\lambda_1 = m$ *and, if* $\operatorname{Re}\lambda_2 < \lambda_1/2$*, the limit in Theorem 3.21 satisfies*

$$\mathbb{E}\,\widehat{V}\widehat{V}' = \sum_{j,k=2}^{q}\biggl(\Bigl(\frac{1}{\lambda_1-\lambda_j}+\frac{1}{\lambda_1-\lambda_k}\Bigr)\frac{u_j'Bu_k}{\lambda_1-\lambda_j-\lambda_k} + \frac{\lambda_1^2-\lambda_j\lambda_k}{(\lambda_1-\lambda_j)(\lambda_1-\lambda_k)}u_j'Du_k\biggr)v_jv_k' \tag{5.4}$$

$$\mathbb{E}\,V\widehat{V}' = \sum_{j,k=2}^{q}\Bigl(\frac{\lambda_1}{(\lambda_1-\lambda_j)(\lambda_1-\lambda_j-\lambda_k)}u_j'Bu_k + \frac{\lambda_1\lambda_j}{\lambda_1-\lambda_j}u_j'Du_k\Bigr)v_jv_k' + \sum_{k=2}^{q}\frac{u_1'Bu_k}{\lambda_1-\lambda_k}v_1v_k' \tag{5.5}$$

Proof. By Lemma 5.4, $\lambda_1 = m$ and $u_1 = a$. Thus, $I - v_1a' = I - v_1u_1' = \sum_{j=2}^{q} v_ju_j'$, and it follows from (3.15) that, with $\phi(s,\lambda) = \int_0^s e^{t\lambda}dt$,

$$\mathbb{E}\,\widehat{V}\widehat{V}' = \sum_{j,k=2}^{q} v_j\int_0^\infty u_j'\Bigl(\phi(s,\lambda_j)B\phi(s,\lambda_k) + e^{s\lambda_j}D + De^{s\lambda_k} - D\Bigr)u_ke^{-\lambda_1 s}\lambda_1\,ds\;v_k'$$

By the definition of $\phi(s,\lambda)$ and changes of order of integration,

$$\begin{aligned}\int_0^\infty \phi(s,\lambda_j)\phi(s,\lambda_k)e^{-\lambda_1 s}\lambda_1\,ds &= \iiint\limits_{0<t,u<s} e^{\lambda_j t+\lambda_k u-\lambda_1 s}\lambda_1\,ds\,du\,dt\\ &= \iint\limits_{0<t<u} e^{\lambda_j t+(\lambda_k-\lambda_1)u}\,du\,dt + \iint\limits_{0<u<t} e^{\lambda_k u+(\lambda_j-\lambda_1)t}\,dt\,du\\ &= \frac{1}{\lambda_1-\lambda_k}\frac{1}{\lambda_1-\lambda_k-\lambda_j} + \frac{1}{\lambda_1-\lambda_j}\frac{1}{\lambda_1-\lambda_j-\lambda_k}.\end{aligned}$$

The integrals of the other terms are easily computed and a summation yields (5.4).

Similarly, since (3.12) and $a = u_1$ imply $u_1'\psi(s,A) = u_1'$ and $u_j'\psi(s,A) = e^{s\lambda_j}u_j'$ for $j \neq 1$, (3.16) yields

$$\begin{aligned}\mathbb{E}\,V\widehat{V}' = \sum_{j,k=2}^{q} v_j\int_0^\infty u_j'e^{s\lambda_j}\bigl(B\phi(s,\lambda_k)+\lambda_jD\bigr)u_ke^{-\lambda_1 s}\lambda_1\,ds\;v_k'\\ + \sum_{k=2}^{q} v_1\int_0^\infty u_1'\bigl(B\phi(s,\lambda_k)+\lambda_1D\bigr)u_ke^{-\lambda_1 s}\lambda_1\,ds\;v_k'\end{aligned}$$

Again, the integrals are easily computed, and (5.5) follows; note that $u_1'Du_k = v_1'u_k = 0$ for $k \neq 1$. □

Remark 5.7. These results can be extended to the covariances in the process limits. For example, under the assumptions of Lemma 5.3(iii), an argument as in the proofs of Lemmas 5.1 and 5.3 shows that in (3.19)

$$\mathbb{E}\, V(x)V(y)' = \sum_{j,k\geq 2} x^{1-\lambda_k/\lambda_1} y^{\lambda_k/\lambda_1} \frac{\lambda_1\, u_j' D u_k}{\lambda_1 - \lambda_j - \lambda_k} w_j w_k', \qquad 0 \leq x \leq y.$$

Similarly, under the assumptions of Lemma 5.4 and with $\operatorname{Re}\lambda_2 < \frac{1}{2}\lambda_1$, we have in (3.19), by Corollary 3.23 with $b := m^{-1}a$ and (3.17),

$$\mathbb{E}\, V(x)V(y)' = mx\Sigma_I (y/x)^{m^{-1}A'}, \qquad 0 < x \leq y. \tag{5.6}$$

6. Relations to previously known results

Large parts of Theorem 3.1 are known since Athreya's thesis, at least in the irreducible case: the a.s. convergence is in [4] and [9, Theorem V.7.2], and the limits in Corollary 3.8 are proved in [5], [6], see also [9, §V.8]. Our results give more explicit formulas for the asymptotic variances and covariances, and the extension to stochastic processes. The independence between the limit processes in (i) and (ii) seems new too.

Also Theorems 3.16–3.19 for urn models are basically due to Athreya and Karlin [8], see also [9, §V.9], but it is not evident how to obtain explicit formulas for asymptotic variances from their paper.

One of the purposes of this paper is to draw attention to the embedding method in [8], adding some details and making the results simpler to apply. In our opinion, this method has been neglected for too long. Several authors have, however, derived similar results for Pólya urns, in more or less general situations, either by calculating moments by recursion formulas or by martingale methods. (Of course, the embedding method uses martingales too, for the branching process. It is thus not really a question of using martingales or not; the main difference is rather whether to use discrete time martingales directly or to first randomize the splitting times by the continuous time branching process and then use continuous time martingales.) It seems that the (discrete time) martingale methods works fine when the number of added balls is fixed, but the extra randomization in the embedding method makes it much easier to handle the general case.

Some important papers with general limit theorems for urn processes that contain special cases of our results are the following. (Our description is brief; for exact conditions and results, see the cited papers. See also further references in these papers.)

Bagchi and Pal [10] gave, using the method of moments, limit theorems when $q = 2$ and the number of added balls is fixed, see Example 7.3 below. Gouet [20] gave functional limit theorems in the same case, using a martingale central limit theorem.

Smythe [51] used martingale methods to establish asymptotic normality of X_n and joint asymptotic normality of X_n and N_n (thus inspiring our Theorem 3.21) when $\lambda_2 < \frac{1}{2}\lambda_1$, allowing removals (in the tenable case) and assuming that the expected number of added balls does not depend on the type of the drawn ball, i.e. that the column sums of A are the same, and thus equal to λ_1. (He also assumed some technical simplifications, for example that A is diagonalizable). However, no general variance formulas were given, although some examples are given and it is stated that asymptotic variances and covariances in principle can be computed more generally. (See also the special case in [44] where asymptotic variances are given.) It is noteworthy that only the expected number of added balls is assumed constant in [51]; most papers applying martingale methods use the stronger assumption that the actual number of added balls is constant.

Bai and Hu [11] used similar martingale methods to establish asymptotic normality of X_n. They assumed that the number of added balls is constant, usually 1, and that $\operatorname{Re}\lambda_2 \leq \frac{1}{2}\lambda_1 = \frac{1}{2}$. A new feature of [11] is that they allow time-dependent transition probabilities (converging to a limit), a case not studied in this paper.

Bai, Hu and Zhang [12] studied the case of two types in more detail, using martingale methods and the Skorohod embedding theorem. Again they allow some time-dependency, and assume that the number of added balls is 1 (with some randomness allowed in the time dependent case), and that $\operatorname{Re}\lambda_2 \leq \frac{1}{2}\lambda_1 = \frac{1}{2}$. They results include functional limit theorems, laws of iterated logarithm, and estimates of rates of convergence.

Remark 6.1. A method that has, as far as we know, *not* yet been used to study this type of urn models is to use a general limit theorem for Markov processes such as [31, Theorem 19.28]. This seems to have the potential of giving comparatively simple proofs of several results in this paper, and could probably be used to attack some of the extensions mentioned in Section 4.

7. Examples and applications to random trees

We give several examples of urn models that illustrate the results above. We concentrate on already studied models and show how several previosly known results follow from our theorems by routine calculations; we encourage the reader to compare the methods. We also give some new results.

Unless otherwise stated, all activities $a_i = 1$, the urn is irreducible, (A1)–(A6) hold, and (essential) extinction is impossible; this can in each case easily be verified using Lemma 2.1. We sometimes omit minor details, such as specifying X_0; similarly, when convenient we shift the indices and start the process with X_1.

Example 7.1. First a trivial example. If $\xi_1, \dots, \xi_q$ are random with the same distribution, the drawn types do not matter and $X_n - X_0$ is a sum of n i.i.d. random vectors. Thus X_n is asymptotically normal by the central limit theorem. In this case, A has rank 1 so $\lambda_2 = 0$ with multiplicity $q-1$, and there are no further eigenvalues. It can be verified that Theorem 3.17 indeed yields the normal limit given by the central limit theorem. Similarly, Theorem 3.24 yields the same result as Donsker's theorem.

If $b = \delta_1 = (1, 0, \dots, 0)'$, then $X_{\tau_b(z)}$ is the vector obtained by summing i.i.d. copies of ξ_1 until the first component is at least z. Corollary 3.11 yields asymptotic normality, as shown in [22], [21].

Example 7.2 (Friedman's urn)**.** A classic example is Friedman's urn [19], [18] (studied already by Bernstein [13], [14] and Savkevitch [49]), where $q = 2$, $\xi_1 = (\alpha, \beta)'$ and $\xi_2 = (\beta, \alpha)'$ for some integers α and β. We assume $\alpha + \beta > 0$. If $\beta = 0$, we have the original Pólya urn [17], [47] which is reducible; (A4), (A5), (A6) fail so our results do not apply, and it is well known that X_n/n converges to a Beta distribution instead of a constant [47], [30].

If $\beta > 0$, our assumptions hold. We have $\lambda_1 = \alpha + \beta$ and $\lambda_2 = \alpha - \beta$, so $\lambda_2 < \frac{1}{2}\lambda_1$ is equivalent to $\alpha < 3\beta$; in this case Theorem 3.17 yields by Lemma 5.3(iii), with $v_1 = \frac{1}{2}(1,1)'$, $v_2 = \frac{1}{2}(1,-1)'$, $u_1 = (1,1)'$, $u_2 = (1,-1)'$,

$$n^{-1/2}\left(X_n - n\frac{\alpha+\beta}{2}\begin{pmatrix}1\\1\end{pmatrix}\right) \xrightarrow{\mathrm{d}} N\left(0, \frac{(\alpha+\beta)(\alpha-\beta)^2}{4(3\beta-\alpha)}\begin{pmatrix}1 & -1\\-1 & 1\end{pmatrix}\right)$$

as shown by Bernstein [13], [14] and Freedman [18].

Similarly, if $\alpha = 3\beta > 0$, Theorem 3.18 and Lemma 5.3(iv) yield [13, 18]

$$(n\ln n)^{-1/2}\left(X_n - n\frac{\alpha+\beta}{2}\begin{pmatrix}1\\1\end{pmatrix}\right) \xrightarrow{\mathrm{d}} N\left(0, \frac{(\alpha-\beta)^2}{4}\begin{pmatrix}1 & -1\\-1 & 1\end{pmatrix}\right).$$

Example 7.3 (general 2-type urn)**.** More generally, consider the case $q = 2$ with nonrandom $\xi_1 = (\alpha, \beta)'$ and $\xi_2 = (\gamma, \delta)'$. We assume that $\beta, \gamma > 0$ so that the urn is irreducible, and that $\lambda_1 > 0$. If $\alpha, \delta \geq -1$, (A1)–(A6) hold and extinction is impossible; by Remark 4.2 we can also allow other negative values for α and δ under suitable conditions, for example in the tenable case.

The case $\alpha + \beta = \gamma + \delta$ has been studied by several authors, including Bernstein [13], [14] and (also for the tenable case) Bagchi and Pal [10], who proved asymptotic normality when $\lambda_2 \leq \lambda_1/2$. Gouet [20] gave functional limit theorems.

We extend their results as follows. We write, for notational convenience, $\kappa := (\alpha + \delta)/2$, $\varepsilon := (\alpha - \delta)/2$, $\rho := \sqrt{\varepsilon^2 + \beta\gamma} > 0$. Thus

$$A = \begin{pmatrix}\alpha & \gamma\\ \beta & \delta\end{pmatrix} = \begin{pmatrix}\kappa+\varepsilon & \gamma\\ \beta & \kappa-\varepsilon\end{pmatrix}.$$

Simple calculations yield the eigenvalues $\lambda_\pm = \kappa \pm \rho$ (the indices $+, -$ are more convenient than $1, 2$ in this example) and the dual bases of eigenvectors (v_+, v_-) and (u_+, u_-) with

$$v_\pm = c_\pm^{-1}\begin{pmatrix}\varepsilon\pm\rho\\ \beta\end{pmatrix}, \qquad u_\pm = \frac{c_\pm}{(\varepsilon\pm\rho)^2+\beta\gamma}\begin{pmatrix}\varepsilon\pm\rho\\ \gamma\end{pmatrix},$$

where $c_\pm$ are normalization constants. We take $c_+ = \beta + \rho + \varepsilon$ so that $a \cdot v_+ = 1$, and choose $c_- = 1$.

If $\lambda_- < \frac{1}{2}\lambda_+$, or, equivalently, $\kappa < 3\rho$, Theorem 3.17 yields $n^{-1/2}(X_n - n(\kappa+\rho)v_+) \xrightarrow{\mathrm{d}} N(0,\Sigma)$, where Lemma 5.3(iii) yields

$$\Sigma = \frac{\lambda_+}{\lambda_+ - 2\lambda_-}(u_-'Du_-)ww',$$

with $w := \lambda_- v_- - \lambda_+(a\cdot v_-)v_+$. By elementary calculations, using $(\rho-\varepsilon)^2 + \beta\gamma = 2\rho(\rho-\varepsilon)$ and $(\rho-\varepsilon)(\beta+\rho+\varepsilon) = \beta(\gamma+\rho-\varepsilon)$, we find

$$u_-'Du_- = \frac{\gamma}{4\rho^2(\rho-\varepsilon)} \qquad \text{and} \qquad w = \frac{2\rho\beta}{\beta+\rho+\varepsilon}\begin{pmatrix}\gamma-\alpha\\ \delta-\beta\end{pmatrix}$$

and

$$\Sigma = \frac{(\kappa+\rho)\beta\gamma}{(3\rho-\kappa)(\beta+\rho+\varepsilon)(\gamma+\rho-\varepsilon)}\begin{pmatrix}\gamma-\alpha\\ \delta-\beta\end{pmatrix}\begin{pmatrix}\gamma-\alpha, & \delta-\beta\end{pmatrix}. \tag{7.1}$$

(It is easy to see that X_{n1} and X_{n2} are linearly dependent, which explains why Σ has rank 1. It is thus sufficient to study only X_{n1}.)

In the special case $\alpha+\beta = \gamma+\delta = m$ studied in [10] this simplifies: now $\lambda_+ = \kappa+\rho = m$, $\rho = (\beta+\gamma)/2$ and $3\rho-\kappa = m+2\rho-2\kappa = m+2(\gamma-\alpha)$, so

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

in accordance with [10].

When $\kappa = 3\rho$, $\lambda_- = \frac{1}{2}\lambda_+$ and Theorem 3.18 yields $(n\ln n)^{-1/2}\bigl(X_n - n(\kappa+\rho)v_+\bigr) \xrightarrow{\mathrm{d}} N(0,\Sigma)$, where by Lemma 5.3(iv) and the calculations above

$$\Sigma = \frac{\beta\gamma}{(\beta+\rho+\varepsilon)(\gamma+\rho-\varepsilon)}\begin{pmatrix}\gamma-\alpha\\ \delta-\beta\end{pmatrix}\begin{pmatrix}\gamma-\alpha, & \delta-\beta\end{pmatrix}. \tag{7.2}$$

In the special case $\alpha+\beta = \gamma+\delta$ (when $\alpha = \beta+2\gamma$, $\delta = 2\beta+\gamma$) this simplifies to $\beta\gamma\left(\begin{smallmatrix}1 & -1\\ -1 & 1\end{smallmatrix}\right)$ as given by [10].

Theorem 3.24 yields functional convergence in $D[0,\infty)$ to a Gaussian process when $\kappa \le 3\rho$. (In the special case $\alpha+\beta = \gamma+\delta$, this was proved by Gouet [20].) If $\kappa < 3\rho$, then

$$n^{-1/2}\bigl(X_{\lfloor xn\rfloor} - xn(\kappa+\rho)v_+\bigr) \xrightarrow{\mathrm{d}} V(x),$$

where $\mathbb{E}\,V(x) = 0$ and, by Remark 5.7, with Σ given in (7.1),

$$\mathbb{E}\,V(x)V'(y) = x^{1-\lambda_-/\lambda_+}y^{\lambda_-/\lambda_+}\Sigma, \qquad 0 \le x \le y.$$

If $\kappa = 3\rho$, then, instead,

$$(\ln n)^{-1/2}n^{-x/2}\bigl(X_{\lfloor n^x\rfloor} - n^x(\kappa+\rho)v_+\bigr) \xrightarrow{\mathrm{d}} V(x),$$

where $\mathbb{E}\,V(x) = 0$ and, with Σ given in (7.2),

$$\mathbb{E}\,V(x)V'(y) = x\Sigma, \qquad 0 \le x \le y.$$

Example 7.4 (randomized play-the-winner)**.** The calculations in Example 7.3 can be extended to random ξ_1 and ξ_2. We consider for simplicity only the case when exactly one ball is added each time. This gives the randomized play-the-winner rule for clinical trials introduced by Wei and Durham [53], see also [54]: If the drawn ball has type i ($i = 1$ or 2), we add a ball with the same type with probability p_i and a ball with the opposite type with probability $q_i = 1 - p_i$. Here p_1 and p_2 are given numbers with $0 \le p_i < 1$. We have

$$A = \begin{pmatrix} p_1 & q_2 \\ q_1 & p_2 \end{pmatrix}.$$

It is easily seen that, see Lemma 5.4, $\lambda_1 = 1$, $\lambda_2 = p_1 + p_2 - 1 = p_1 - q_2$,

$$u_1 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \quad v_1 = \frac{1}{q_1+q_2}\begin{pmatrix} q_2 \\ q_1 \end{pmatrix}, \quad u_2 = \begin{pmatrix} q_1 \\ -q_2 \end{pmatrix}, \quad v_2 = \frac{1}{q_1+q_2}\begin{pmatrix} 1 \\ -1 \end{pmatrix}.$$

Lemma 5.5 yields

$$B = D = \frac{1}{q_1+q_2}\begin{pmatrix} q_2 & 0 \\ 0 & q_1 \end{pmatrix}.$$

When $\lambda_2 < 1/2$, Theorem 3.17 applies with, by Lemmas 5.4 and 5.3(i),

$$\Sigma = \Sigma_I = \frac{u_2' B u_2}{1-2\lambda_2} v_2 v_2' = \frac{q_1 q_2}{(1-2\lambda_2)(q_1+q_2)^2}\begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix}.$$

(Note that $q_1 + q_2 = 1 - \lambda_2$.) Moreover, by Theorem 3.21(i), we have joint asymptotic normality of the numbers of balls of different types after n draws and the numbers of drawn balls of different types. Considering, as we may, only type 1, we have

$$n^{-1/2}\bigl(X_{n1} - nq_2/(q_1+q_2),\, N_{n1} - nq_2/(q_1+q_2)\bigr) \xrightarrow{\mathrm{d}} N(0, \Sigma_1),$$

where Lemma 5.6 easily gives

$$\Sigma_1 = \frac{q_1 q_2}{(1-2\lambda_2)(q_1+q_2)^2}\begin{pmatrix} 1 & 1+2\lambda_2 \\ 1+2\lambda_2 & 3+2\lambda_2 \end{pmatrix}.$$

Similarly, when $\lambda_2 = 1/2$, Theorem 3.18 applies with $d = 0$ and, by Lemmas 5.4 and 5.3(i),

$$\Sigma = \Sigma_{II,d} = \Sigma_{II} = \frac{q_1 q_2}{(q_1+q_2)^2}\begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix} = 4q_1 q_2 \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix}.$$

Furthermore, Theorem 3.21(ii) applies, and $\widehat{V} = A^{-1} V = 2V$.

For earlier proofs of these results and some extensions, see [52], [51], [11], [12].

Example 7.5 (random 2-3 trees)**.** Bagchi and Pal [10] applied their general result to random 2-3 trees. In such trees, all internal nodes have 2 or 3 children, and all external nodes (leaves) are at the same distance from the root. Keys are associated either with the leaves or with the internal nodes [10], [2]. We define, following [10], the type of an internal node to be W if it has 2 children and B if it has 3; an external node has the same type as its parent. When the tree is grown randomly, a new external node is inserted adjacent to a randomly chosen old one; this either transforms a W-type node at the lowest level to B, or splits a B-type node at the lowest level into two nodes of type W (possibly inducing further splits higher up). Bagchi and Pal [10] study the types of the external nodes as an urn process, with $\xi_1 = (-2,3)'$ and $\xi_2 = (4,-3)'$. This does not satisfy (1.2) so, as noted in [10], the Athreya–Karlin embedding is not immediately available. However, this example is of the "tenable" type described in Remark 4.2, where we can use the superball trick. In this case, the superball method is very natural; it means that we consider the lowest level internal nodes instead of the external nodes. This "internal" urn model for 2-3 trees was one of the examples considered by Aldous, Flannery and Palacios [2], who noted that the Athreya–Karlin embedding works.

For the internal version, we have the urn model with $\xi_1 = (-1,1)'$ and $\xi_2 = (2,-1)'$ and the activities $a_1 = 2$, $a_2 = 3$; hence $A = \left(\begin{smallmatrix} -2 & 6 \\ 2 & -3 \end{smallmatrix}\right)$. Straightforward calculations yield $\lambda_1 = 1$, $\lambda_2 = -6$ and, by Theorem 3.17 and Lemma 5.3(iii), for example $n^{-1/2}(W_n - \frac{2}{7}n) \xrightarrow{\mathrm{d}} N(0, \frac{108}{637})$, if W_n is the number of W-type internal nodes in the lowest level when we have n nodes. For w_n, the number of W-type external nodes, we have $w_n = 2W_n$ and thus

$$n^{-1/2}(w_n - \tfrac{4}{7}n) \xrightarrow{\mathrm{d}} N(0, \tfrac{432}{637}), \tag{7.3}$$

as shown by other methods in [10].

Alternatively, we can obtain this directly by using Theorem 3.17 on the external version in [10], although (A1) is not satisfied; as remarked in Remark 4.2, this is allowed for tenable urns (and for some other urns too). In this direct approach, $A = \left(\begin{smallmatrix} -2 & 4 \\ 3 & -3 \end{smallmatrix}\right)$, and (7.3) follows by simple computations using Lemma 5.3(iii), or directly by (7.1). Note that, as always with the superball trick, A differs for the two versions, but the eigenvalues are the same, see Remark 4.2.

Example 7.6 (random recursive trees)**.** Mahmoud and Smythe [43] used a generalized Pólya urn to study random recursive trees and obtained the asymptotic normal distribution of the nodes of outdegrees 0, 1 and 2. They indicated that the results in principle extend to higher degrees; we can now do this.

The distribution of outdegrees is the same as the distribution of types in a generalized Pólya urn with *infinitely* many types $\{0,1,2,\dots\}$ and the rule that if a ball with type i is drawn, it is removed and replaced by a ball of type $i+1$ and a ball of type 0, see [43]. In our notation, $\xi_{ij} = -\delta_{ij} + \delta_{0j} + \delta_{i+1,j}$. Our theorems assume that the number of types is finite, but luckily we can in this application truncate and lump all high degrees together. Thus, let $M \geq 1$ be an integer and use the types $\{0,1,\dots,M\}$ only (thus $q = M+1$), where now type M represents all outdegrees $\geq M$. The replacement vectors ξ_i are as in the infinite model when $i < M$, while now $\xi_M = (1,0,\dots,0)'$.

Exactly one ball is added each time, so $\lambda_1 = 1$ and $u_1 = a = (1, 1, \dots, 1)'$ by Lemma 5.4. It is easily verified that $v_1 = (1/2, 1/4, \dots, 2^{-M}, 2^{-M})'$, i.e. $v_{1i} = 2^{-i-1}$ for $0 \le i < M$ and $v_{1M} = 2^{-M}$. In particular, Theorem 3.16 shows that $X_{ni}/n \overset{\text{a.s.}}{\to} 2^{-i-1}$ for every $i \ge 0$ (by taking $M > i$); the weaker statement $X_{ni}/n \overset{\text{p}}{\to} 2^{-i-1}$ was shown by Meir and Moon [45].

It can be shown, see [29], that A has besides $\lambda_1 = 1$ only the eigenvalue -1 (with multiplicity $q - 1 = M$). (Moreover, $d_{-1} = M - 1$, so A is not diagonalizable when $M \ge 2$.) Since thus $\lambda_2 = -1$, Theorem 3.17 applies for every M, and the vector $\big(n^{-1/2}(X_{ni} - 2^{-i-1}n)\big)_{i=0}^{M-1}$ converges in distribution to a Gaussian vector. Since M is arbitrary, this is the same as convergence of the infinite vector $\big(n^{-1/2}(X_{ni} - 2^{-i-1}n)\big)_{i=0}^{\infty}$ in $\mathbb{R}^\infty$, see [15, p. 19]. In other words, $n^{-1/2}(X_{ni} - 2^{-i-1}n) \overset{\text{d}}{\to} V_i$, jointly for all $i \ge 0$, as $n \to \infty$, where the V_i are jointly Gaussian variables with means $\mathbb{E}\, V_i = 0$. The (co)variances $\Sigma_{jk} := \operatorname{Cov}(V_j, V_k)$ are calculated in [29] using Lemma 5.1.

Similarly, Theorem 3.24 yields a functional limit theorem: $n^{-1/2}(X_{\lfloor xn \rfloor, i} - 2^{-i-1}xn) \overset{\text{d}}{\to} V_i(x)$ in $D[0, \infty)$, where the $V_i(x)$ are continuous Gaussian processes with $\mathbb{E}\, V_i(x) = 0$. Again, see [29] for covariances.

Example 7.7 (random plane recursive trees)**.** Mahmoud, Smythe and Szymański [44] studied random plane recursive trees and obtained (among other results) the asymptotic normal distribution of the number of nodes of outdegrees 0, 1 and 2. The outdegrees can be modelled using a generalized Pólya urn with infinitely many colours as in Example 7.6; the ξ_{ij} are the same, but now the activity $a_i = i + 1$. In this case it is advantageous to use the reverse of the superball trick: we replace each ball of type i by $i + 1$ balls of the same type. (The new balls can be interpreted as external vertices as in [44].) This yields a new generalized Pólya urn with infinitely many types, all activities 1, and the transitions given by

$$\xi_{ij} = -(j+1)\delta_{ij} + \delta_{0j} + (j+1)\delta_{i+1,j} = -(i+1)\delta_{ij} + \delta_{0j} + (i+2)\delta_{i+1,j}.$$

Again we truncate and use the $M+1$ types $0, \dots, M$ only, with ξ_{Mj} changed to $\delta_{0j} + \delta_{Mj}$. (Note that such truncation does not work in the original urn model representing internal nodes.)

In this case the eigenvalues are $\lambda_1 = 2$ and $-1, -2, \dots, -M$, see [29]. Theorem 3.17 applies for every M, and extends the joint asymptotic normality found by Mahmoud, Smythe and Szymański [44] to all degrees. Theorem 3.24 yields a functional limit theorem. The (co)variances are computed, using Lemmas 5.4 and 5.3, in [29].

Example 7.8 (rotations in a binary tree)**.** Mahmoud [39] modelled rotations in the construction of a fringe-balanced binary tree by a generalized Pólya urn with three types, with $X_0 = (2, 0, 0)'$ and $\xi_1 = (-2, 1, 2)'$, $\xi_2 = \xi_3 = (4, -1, -2)'$. The number R_n of rotations in the n first insertions in the binary tree then equals the number of times a ball of type 3 is drawn, i.e. $R_n = N_{n3}$.

Note that $\xi_{23}, \xi_{32} < 0$, so (1.1) is violated.

Nevertheless, Mahmoud [39] observed that the proof of asymptotic normality in Smythe [51] holds for this urn too, and after finding exact formulas for the mean and variance he obtained

$$n^{-1/2}\big(R_n - \tfrac{2}{7}n\big) \xrightarrow{\mathrm{d}} N\big(0, \tfrac{66}{637}\big). \tag{7.4}$$

Although (1.1), and thus (A1), does not hold, we can derive this result from our theorems in several ways; since these methods may be useful in other applications where some ξ_{ij} are negative, we sketch three different approaches, leaving simple calculations to the reader. (It is instructive to compare the different calculations leading to the same result.)

First, note that in this urn, X_{n1} is even and $X_{n3} = 2X_{n2}$, which guarantees that we are never required to remove balls that do not exist. Moreover, $\lambda_1 = 1$ with eigenvectors $u_1 = (1,1,1)'$ and $v_1 = \frac{1}{7}(4,1,2)'$. Thus, as asserted in Remark 4.2, our theorems hold for this urn too, and (7.4) follows by Theorem 3.21 and Lemma 5.6.

Secondly, since $\xi_2 = \xi_3$, we may combine types 2 and 3 and consider the urn with two types and $\xi_1 = (-2,3)'$, $\xi_2 = (4,-3)'$; to obtain R_n we add a dummy ball with probability 2/3 each time we draw a ball of type 2. This is a tenable urn (the same as in Example 7.5), and the result can be obtained by applying Corollary 3.11 and Lemma 5.3(i) to this urn (with dummy balls), stopping when $X_{n1} + X_{n2} = n + 2$.

Thirdly, consider the same 2-type urn again, but instead of adding dummy balls at random, observe that given the number N_{n2} of type 2 draws, $R_n \sim \mathrm{Bi}(N_{n2}, 2/3)$. By Theorem 3.21, or by Example 7.5 and $N_{n2} = \frac{1}{6}\big(3n - X_{n2}\big) = \frac{1}{6}\big(2n + w_n - w_0\big)$,

$$n^{-1/2}\big(N_{n2} - \tfrac{3}{7}n\big) \xrightarrow{\mathrm{d}} N(0, \tfrac{12}{637}). \tag{7.5}$$

In particular, $N_{n2}/n \xrightarrow{\mathrm{p}} 3/7$. Hence, the central limit theorem for the binomial distribution implies $n^{-1/2}\big(R_n - \frac{2}{3}N_{n2}\big) \xrightarrow{\mathrm{d}} N(0, \frac{3}{7}\cdot\frac{2}{3}\cdot\frac{1}{3})$; moreover, this holds jointly with (7.5), with independent limits. Thus

$$n^{-1/2}\big(R_n - \tfrac{2}{7}n\big) = n^{-1/2}\big(R_n - \tfrac{2}{3}N_{n2}\big) + \tfrac{2}{3}n^{-1/2}\big(N_{n2} - \tfrac{3}{7}n\big) \xrightarrow{\mathrm{d}} N\big(0, \tfrac{2}{21} + \big(\tfrac{2}{3}\big)^2 \tfrac{12}{637}\big).$$

This also shows that of the variance in (7.4), only a fraction $\big(\frac{2}{3}\big)^2 \frac{12}{66} = \frac{8}{99}$ comes from the random variation in the urn, i.e. from the shape of the tree.

In the second and third methods, we reduce to the tenable urn for external nodes in Example 7.5. We can replace this urn by the urn for internal nodes in Example 7.5, thus reducing the problem further to an urn that satisfies (A1). This is equivalent to reducing the original 3-type urn by an extension of the superball trick, where the superballs may combine balls of different types. In this case we have two types of superballs: the first represents two balls of type 1, while the second represents 3 balls, 1 of type 2 and 2 of type 3. (This yields two further ways of deriving (7.4).)

Example 7.9 (random m-ary search tree)**.** In an m-ary search tree, where $m \geq 2$ is a fixed integer, a node may contain up to $m-1$ keys. The tree is constructed recursively, starting with an empty root node. Incoming keys are added to the root node until it is full; it then get m daughters, initially empty, and further keys are passed on to one of the daughters, where the procedure repeats. See e.g. [38, Chapter 3].

Let us say that a node containing i keys has type i. With random input, the number of nodes of different types is modelled by a generalized Pólya urn with m types $0, \dots, m-1$. A ball of type i with $i < m-2$ has activity $i+1$; if drawn, it is removed and replaced by a ball of type $i+1$. A ball of type $m-2$ has activity $m-1$; if drawn, it is removed and replaced by a ball of type $m-1$ and m balls of type 0. A ball of type $m-1$ has activity 0. (Since balls of type $m-1$ are dead, we can ignore them when studying the other types.)

Alternatively, we can study external vertices; each (internal) vertex of type $i \leq m-2$ has $i+1$ external vertices, which we label with the same type i. The external vertices evolve as an urn with $m-1$ types $0, \dots, m-2$, all activities 1, and the replacement rules $\xi_i = -(i+1)\delta_i + (i+2)\delta_{i+1}$, $0 \leq i \leq m-3$, and $\xi_{m-2} = -(m-1)\delta_{m-2} + m\delta_0$. For example, for $m = 4$ the external version has the matrix

$$A = \begin{pmatrix} -1 & 0 & 4 \\ 2 & -2 & 0 \\ 0 & 3 & -3 \end{pmatrix}$$

For the external version, it is easily seen that A has the characteristic polynomial $\phi_m(\lambda) := \prod_{i=1}^{m-1}(\lambda + i) - m!$; the largest real root is $\lambda_1 = 1$, see Lemma 5.4. (For the internal version, we have $\lambda\phi_m(\lambda)$ with an additional root $\lambda = 0$.)

A detailed study [42] shows that $\operatorname{Re}\lambda_2 \to 1 = \lambda_1$ as $m \to \infty$, and that $\operatorname{Re}\lambda_2 < \frac{1}{2}$ for $m \geq 26$, but $\operatorname{Re}\lambda_2 > \frac{1}{2}$ for $m > 26$. Hence, the numbers of nodes of different types have an asymptotic normal distribution when $m \leq 26$, but, as rigorously shown by Chern and Hwang [16], not for larger m. This has earlier been shown by other methods [42], [36], [16], and by urns as here by [41].

Example 7.10 (a branched random walk)**.** Let G be a finite group and ξ a random element of G, with some distribution μ. We define an urn process where the types are the elements of G and $\xi_{gh} = \delta_{g\xi,h}$. We thus draw a ball, replace it and add a new ball with a type shifted according to μ. This is a special type of a branching random walk on G.

Since exactly one ball is added each time, $\lambda_1 = 1$. By the symmetry, $u_1 = (1, \dots, 1)'$ and $v_1 = q^{-1}(1, \dots, 1)'$. The matrix A operates by convolution on G: $Av = v*\mu$ for $v \in C^q = \ell^2(G)$. Hence, if G is commutative, the characters of G are eigenvectors of A and the eigenvalues are the Fourier coefficients of μ: $A\chi = \hat{\mu}(\chi)\chi$, $\chi \in \widehat{G}$. In particular, we see that $n^{-1/2}(X_n - nv_1)$ converges to a Gaussian limit $N(0, \Sigma)$ if and only if $\operatorname{Re}\hat{\mu}(\chi) < 1/2$ for all

$\chi \neq 1$. Lemma 5.5 yields $B = q^{-1}I$ (for any μ). Hence, Lemmas 5.4 and 5.3(i) together with the orthogonality of the characters yield

$$\Sigma = \Sigma_I = q^{-2} \sum_{\chi \neq 1} \frac{1}{1 - 2 \operatorname{Re} \hat{\mu}(\chi)} \chi \overline{\chi}'.$$

For a concrete example, let G be the cyclic group $\mathbb{Z}_q$ and suppose that only nearest-neighbour shifts are allowed, i.e. μ is supported on ± 1. (For example, we may always shift one step forward, or make a symmetric random choice each time.) Then $\operatorname{Re} \hat{\mu}(\chi_k) = \cos(2\pi k/q)$, $k = 0, \dots, q-1$, and $\operatorname{Re} \lambda_2 = \cos(2\pi/q)$. Hence, if $q \leq 5$, $\operatorname{Re} \lambda_2 < 1/2$ and X_n is asymptotically normal with variance of the order n; if $q = 6$, X_n is still asymptotically normal (but more degenerate) but the variance is of order $n \log n$, and if $q > 6$, the variance is of larger order and X_n is not asymptotically normal.

For non-commutative G, we obtain similar results by considering the irreducible representations of G.

These results were proved in [25] by a different method (moment calculations). Moreover, [25] treats also *infinite* compact groups, obtaining the same results there. This suggests that the results in this paper may have generalizations to infinite sets of types, see Remark 4.1.

8. A lemma

We will later use a lemma on joint convergence in distribution. The lemma is a simple exercise in measure theory, but since we do not know any good reference, and the notation makes it look more complicated than it really is, we give a detailed statement and proof. We begin with a simpler version.

Lemma 8.1. *Suppose that (η_n, ζ_n) are pairs of random variables with values in $\mathcal{S}_1 \times \mathcal{S}_2$ for some separable metric spaces $\mathcal{S}_1$ and $\mathcal{S}_2$, and that η and ζ are random variables with values in $\mathcal{S}_1$ and $\mathcal{S}_2$, respectively. Suppose further:*

(i) *$\eta_n \xrightarrow{\mathrm{d}} \eta$ as $n \to \infty$;*
(ii) *for every measurable set $A \subseteq \mathcal{S}_1$ such that $\liminf_{n\to\infty} \mathbb{P}(\eta_n \in A) > 0$, it holds that, conditioned on $\eta_n \in A$, $\zeta_n \xrightarrow{\mathrm{d}} \zeta$.*

Then we have joint convergence $(\eta_n, \zeta_n) \xrightarrow{\mathrm{d}} (\eta', \zeta')$ as $n \to \infty$, with η' and ζ' independent copies of η and ζ, respectively.

Proof. Suppose that A and B are measurable sets in $\mathcal{S}_1$ and $\mathcal{S}_2$ with $\mathbb{P}(\eta \in \partial A) = \mathbb{P}(\zeta \in \partial B) = 0$. Then, by (i), $\mathbb{P}(\eta_n \in A) \to \mathbb{P}(\eta \in A)$, see e.g. [15, Theorem 2.1]. If $\mathbb{P}(\eta \in A) > 0$, we thus have by (ii) that, conditioned on $\eta_n \in A$, $\zeta_n \xrightarrow{\mathrm{d}} \zeta$, and thus $\mathbb{P}(\zeta_n \in B \mid \eta_n \in A) \to \mathbb{P}(\zeta \in B)$. Consequently, when $\mathbb{P}(\eta \in A) > 0$,

$$\begin{aligned} \mathbb{P}\big((\eta_n, \zeta_n) \in A \times B\big) &= \mathbb{P}(\zeta_n \in B \mid \eta_n \in A)\, \mathbb{P}(\eta_n \in A) \\ &\to \mathbb{P}(\zeta \in B)\, \mathbb{P}(\eta \in A) = \mathbb{P}\big((\eta', \zeta') \in A \times B\big). \end{aligned} \tag{8.1}$$

The same holds trivially if $\mathbb{P}(\eta \in A) = 0$, and $(\eta_n, \zeta_n) \xrightarrow{\mathrm{d}} (\eta', \zeta')$ follows by [15, Theorem 3.1]. □

We will need the following extension. (The index n may be replaced by a continuous parameter t, since it suffices to consider sequences $t_n \to \infty$.)

Lemma 8.2. *Suppose that η_n, ζ_n, η, ζ, $\mathcal{S}_1$ and $\mathcal{S}_2$ are as in Lemma 8.1. Let further $E_1 \subseteq E_2 \subseteq \dots$ be an increasing sequence of measurable subsets of $\mathcal{S}_1$, let $E = \bigcup_1^\infty E_m$, and suppose that:*

(i) $\eta_n \xrightarrow{\mathrm{d}} \eta$ *as* $n \to \infty$;
(ii) $\mathbb{P}(\eta \in \partial E_m) = 0$ *for every* $m = 1, 2, \dots$;
(iii) $\mathbb{P}(\eta_n \in E) \to \mathbb{P}(\eta \in E) > 0$ *as* $n \to \infty$;
(iv) *for every $m = 1, 2, \dots$ and every measurable set $A \subseteq E_m$ such that $\liminf_{n\to\infty} \mathbb{P}(\eta_n \in A) > 0$, it holds that, conditioned on $\eta_n \in A$, $\zeta_n \xrightarrow{\mathrm{d}} \zeta$.*

Then, conditioned on $\eta_n \in E$, we have $\zeta_n \xrightarrow{\mathrm{d}} \zeta$ and joint convergence

$$\mathcal{L}\big((\eta_n, \zeta_n) \mid \eta_n \in E\big) \xrightarrow{\mathrm{d}} \mathcal{L}\big((\eta', \zeta') \mid \eta' \in E\big),$$

with η' and ζ' independent copies of η and ζ, respectively.

Proof. Suppose again that A and B are measurable sets in $\mathcal{S}_1$ and $\mathcal{S}_2$ with $\mathbb{P}(\eta \in \partial A) = \mathbb{P}(\zeta \in \partial B) = 0$. Let $\varepsilon > 0$ and choose m such that $\mathbb{P}(\eta \in E \setminus E_m) < \varepsilon$. By (i), (ii) and (iii) we have

$$\begin{aligned}\mathbb{P}(\eta_n \in E \setminus E_m) &= \mathbb{P}(\eta_n \in E) - \mathbb{P}(\eta_n \in E_m)\\ &\to \mathbb{P}(\eta \in E) - \mathbb{P}(\eta \in E_m) = \mathbb{P}(\eta \in E \setminus E_m),\end{aligned}$$

and thus $\mathbb{P}(\eta_n \in E \setminus E_m) < \varepsilon$ too for large n.

By (ii), $\mathbb{P}(\eta \in \partial(A \cap E_m)) \le \mathbb{P}(\eta \in \partial A) + \mathbb{P}(\eta \in \partial E_m) = 0$. Hence, (iv) implies as in (8.1)

$$\mathbb{P}\big((\eta_n, \zeta_n) \in (A \cap E_m) \times B\big) \to \mathbb{P}\big((\eta', \zeta') \in (A \cap E_m) \times B\big)$$

and thus

$$\begin{aligned}&\big|\mathbb{P}\big((\eta_n, \zeta_n) \in (A \cap E) \times B\big) - \mathbb{P}\big((\eta', \zeta') \in (A \cap E) \times B\big)\big|\\ &\quad= \big|\mathbb{P}\big((\eta_n, \zeta_n) \in (A \cap (E \setminus E_m)) \times B\big)\\ &\qquad+ \mathbb{P}\big((\eta_n, \zeta_n) \in (A \cap E_m) \times B\big) - \mathbb{P}\big((\eta', \zeta') \in (A \cap E_m) \times B\big)\\ &\qquad- \mathbb{P}\big((\eta', \zeta') \in (A \cap (E \setminus E_m)) \times B\big)\big|\\ &\quad\le \mathbb{P}(\eta_n \in E \setminus E_m)\\ &\qquad+ \big|\mathbb{P}\big((\eta_n, \zeta_n) \in (A \cap E_m) \times B\big) - \mathbb{P}\big((\eta', \zeta') \in (A \cap E_m) \times B\big)\big|\\ &\qquad+ \mathbb{P}(\eta \in E \setminus E_m)\\ &\quad< 3\varepsilon,\end{aligned}$$

provided n is large enough. Consequently, as $n \to \infty$,

$$\mathbb{P}\big((\eta_n, \zeta_n) \in (A \cap E) \times B\big) \to \mathbb{P}\big((\eta', \zeta') \in (A \cap E) \times B\big).$$

Dividing by $\mathbb{P}(\eta_n \in E) \to \mathbb{P}(\eta \in E) = \mathbb{P}(\eta' \in E)$, we find

$$\mathbb{P}\big((\eta_n, \zeta_n) \in A \times B \mid \eta_n \in E\big) \to \mathbb{P}\big((\eta', \zeta') \in A \times B \mid \eta' \in E\big),$$

and the result follows by [15, Theorem 3.1]. □

Note that (iii) follows from (i) if $\mathbb{P}(\eta \in \partial E) = 0$. However, this stronger condition does not always hold in our applications.

9. Proof of Theorem 3.1

The proof of Theorem 3.1 is based on martingale theory, in particular a martingale convergence theorem by Jacod and Shiryaev [24]. The theorem uses the *quadratic variation* $[X, X]_t$ of a martingale X defined on $[0, \infty)$, and its bilinear extension $[X, Y]_t$ to two martingales X and Y. For a general definition see e.g. [24] or [48]; for us it will suffice to know that, if X and Y are (real or complex) martingales of finite variation, then

$$[X, Y]_t = \sum_{0<s\le t} \Delta X(s) \Delta Y(s), \tag{9.1}$$

where $\Delta X(s) := X(s) - X(s-)$ is the jump of X at s and, similarly, $\Delta Y(s) := Y(s) - Y(s-)$. (A martingale X is said to be of finite variation if it is so pathwise, i.e. if $t \mapsto X(t)$ a.s. has bounded variation on each finite interval.) The sum in (9.1) is formally uncountable, but in reality countable since there is only a countable number of jumps; in the applications below, the sum will be finite. (There is some disagreement in the literature on the definition of $[X, Y]$ in the case $X(0)Y(0) \neq 0$; we have chosen the version with $[X, Y]_0 = 0$.) For martingales of infinite variation (such as Brownian motion), (9.1) fails, but we have always the inequality

$$\sum_{0<s\le t} |\Delta X(s)|^2 \le [X, \bar{X}]_t. \tag{9.2}$$

For vector-valued martingales $X = (X_i)_{i=1}^m$ and $Y = (Y_j)_{j=1}^n$, we define the square bracket $[X, Y]$ to be the $m \times n$ matrix $\big([X_i, Y_j]\big)_{i,j}$.

For a real-valued martingale X, the quadratic variation $[X, X]_t$ is a non-negative and non-decreasing process. A real-valued martingale $X(s)$ on $[0, t]$ is an L^2-martingale if and only if $\mathbb{E}[X, X]_t < \infty$ and $\mathbb{E}\,|X(0)|^2 < \infty$, and then

$$\mathbb{E}\,|X(t)|^2 = \mathbb{E}[X, X]_t + \mathbb{E}\,|X(0)|^2. \tag{9.3}$$

(For complex-valued martingales one has to consider $[X, \bar{X}]_t$.) There is also a corresponding bilinear formula, which extends to (real or complex) vector-valued L^2-martingales in the form

$$\mathbb{E}\,X(t)Y'(t) = \mathbb{E}[X, Y]_t + \mathbb{E}\,X(0)Y'(0). \tag{9.4}$$

We will use the following general result based on [24]; see [26] and [27] for similar versions. Again, n may be replaced by a continuous parameter t.

Proposition 9.1. (i) *Assume that for each n, $M_n(x) = (M_{ni}(x))_{i=1}^q$ is a real q-dimensional martingale on $[0,\infty)$ with $M_n(0) = 0$, and that $\Sigma(x)$, $x \geq 0$, is a (non-random) continuous matrix-valued function such that for every fixed $x \geq 0$,*

$$[M_n, M_n]_x \xrightarrow{\text{p}} \Sigma(x) \quad \text{as } n \to \infty, \tag{9.5}$$

$$\sup_n \mathbb{E}\,|M_n(x)|^2 < \infty. \tag{9.6}$$

Then $M_n \xrightarrow{\text{d}} M$ as $n \to \infty$, in $D[0,\infty)$, where M is a continuous q-dimensional Gaussian process with $\mathbb{E}\,M(x) = 0$ and covariances

$$\mathbb{E}\,M(x)M'(y) = \Sigma(x), \quad 0 \leq x \leq y < \infty. \tag{9.7}$$

(ii) *The same holds for complex M_n (with M complex), provided* (9.5) *is supplemented by*

$$[M_n, \overline{M}_n]_x \xrightarrow{\text{p}} \Sigma^\dagger(x) \quad \text{as } n \to \infty, \tag{9.8}$$

for some continuous matrix-valued function $\Sigma^\dagger(x)$, and then further

$$\mathbb{E}\,M(x)\overline{M}'(y) = \Sigma^\dagger(x), \quad 0 \leq x \leq y < \infty.$$

Proof. (i): Note first that (9.5) implies that if $x \leq y$, then $\Sigma(y) - \Sigma(x)$ is positive semidefinite, so there exists a q-dimensional Gaussian process M with independent increments such that $\mathbb{E}\,M(x) = 0$ and $\mathbb{E}\,M(x)M'(y) = \Sigma(x)$, $x \leq y$, see e.g. [24, Theorem II.5.2]. M is a martingale, and it is continuous because each component is a (deterministic) time change of a Brownian motion.

Since, by (9.2) and (9.3),

$$\mathbb{E}\sup_{y\leq x}|\Delta M_{ni}(y)|^2 \leq \mathbb{E}\sum_{y\leq x}|\Delta M_{ni}(y)|^2 \leq \mathbb{E}[M_{ni}, M_{ni}]_x = \mathbb{E}\,|M_{ni}(x)|^2,$$

it follows from (9.6) that, for each fixed $x > 0$, the sequence $\sup_{y\leq x}|\Delta M_{ni}(y)|$ is uniformly integrable for each i, and thus $\sup_{y\leq x}|\Delta M_n(s)|$ is uniformly integrable. The result now follows from [24, Theorem VIII.3.12, (ii)⇒(i)].

(ii): This follows from (i) by considering the real $2q$-dimensional martingales $(\operatorname{Re} M_n, \operatorname{Im} M_n)$. Note that (9.5) for these (with the appropriate right-hand sides) follows from (9.5) and (9.8). □

In order to apply this result to our process $\mathcal{X}(t)$, we have to first define a suitable martingale, and then estimate its quadratic variation. The martingale is a standard one in branching process theory, and the estimates will be derived by standard methods too, although the details will take some time. We proceed with a series of lemmas. (Some of these are known, but included here for completeness and because our conditions are slightly more general than the standard ones; see also Remark 4.2.) We make the definition

$$\mathcal{Y}(t) := e^{-tA}\mathcal{X}(t) \tag{9.9}$$

and begin with a fundamental well-known result, cf. [9, Theorem V.8.1].

Lemma 9.2. *$\mathcal{Y}(t)$ is a martingale for $t \geq 0$. In particular,*

$$\mathbb{E}\,\mathcal{X}(t) = e^{tA}\,\mathbb{E}\,\mathcal{X}(0) \tag{9.10}$$

and thus

$$\mathbb{E}\,\mathcal{X}(t) = O\big(e^{\lambda_1 t}\big). \tag{9.11}$$

Proof. It follows from the definitions of $\mathcal{X}$ and A that

$$\frac{d}{dt}\,\mathbb{E}\,\mathcal{X}_j(t) = \sum_{i=1}^{q} a_i\,\mathbb{E}\,\mathcal{X}_i(t)\,\mathbb{E}\,\xi_{ij} = \big(A\,\mathbb{E}\,\mathcal{X}(t)\big)_j \tag{9.12}$$

and hence $\frac{d}{dt}\,\mathbb{E}\,\mathcal{X}(t) = A\,\mathbb{E}\,\mathcal{X}(t)$. This yields (9.10) by integration, and the martingale property follows from (9.10) and the Markov property. Finally, (2.12) implies (9.11). $\square$

Let $0 < \tau_{i1} < \tau_{i2} < \dots$ denote the times a ball of type i splits, and let $N_i(t) := \#\{k : \tau_{ik} \leq t\}$ be the number of such splits up to time t. Since the martingale $\mathcal{Y}(t) := e^{-tA}\mathcal{X}(t)$ has finite variation, its quadratic variation is by (9.1) given by its jumps

$$\begin{aligned}[]
[\mathcal{Y},\mathcal{Y}]_t &= \sum_{i=1}^{q} \sum_{k:\tau_{ik}\leq t} \Delta\mathcal{Y}(\tau_{ik})\Delta\mathcal{Y}'(\tau_{ik}) \\
&= \sum_{i=1}^{q} \sum_{k:\tau_{ik}\leq t} e^{-\tau_{ik}A}\Delta\mathcal{X}(\tau_{ik})\Delta\mathcal{X}'(\tau_{ik})e^{-\tau_{ik}A'}.
\end{aligned} \tag{9.13}$$

The main part of the proof consists of estimating this sum, and components of it. It will be convenient to state a general lemma for sums of this type, for simplicity considering a single i.

Lemma 9.3. *Fix $i \in \{1,\dots,q\}$. Let f_1 and f_2 be continuous matrix-valued functions defined on $[0,\infty)$ and let g be a matrix-valued function on $\mathbb{R}^q$ such that $\mathbb{E}\,|g(\xi_i)| < \infty$. Suppose further that the dimensions of f_1, g, and f_2 are such that the product $f_1 g f_2$ is defined. Let*

$$Z(t) := \sum_{k:\tau_{ik}\leq t} f_1(\tau_{ik})g\big(\Delta\mathcal{X}(\tau_{ik})\big)f_2(\tau_{ik}) = \int_0^t f_1(s)g\big(\Delta\mathcal{X}(s)\big)f_2(s)\,dN_i(s) \tag{9.14}$$

and

$$\widetilde{Z}(t) := Z(t) - \int_0^t f_1(s)(\mathbb{E}\,g(\xi_i))f_2(s)a_i\mathcal{X}_i(s)\,ds. \tag{9.15}$$

Then $\widetilde{Z}(t)$, $t \geq 0$, is a (matrix-valued) martingale; in particular

$$\mathbb{E}\,Z(t) = \int_0^t f_1(s)\,\mathbb{E}\,g(\xi_i)f_2(s)a_i\,\mathbb{E}\,\mathcal{X}_i(s)\,ds. \tag{9.16}$$

Proof. In the special case when $f_1 = f_2 = g = 1$, $Z(t) = N_i(t)$ and $\widetilde{Z}(t) = \widetilde{N}_i(t) := N_i(t) - \int_0^t a_i \mathcal{X}_i(s)\,ds$. The fact that $\widetilde{N}_i(t)$ is a martingale is a well-known simple consequence of the assumption that the balls have independent exponential lifetimes. The present extension can be proved in the same way, because $\Delta\mathcal{X}(\tau_{ik})$ is independent of the previous history; for example, a straightforward calculation shows that $\widetilde{Z}(t\wedge\tau_n) - \widetilde{Z}(t\wedge\tau_{n-1})$ (with $\tau_0 := 0$) is a martingale for each n, and the result follows by summing over n. We omit the details. □

Note that (9.16) with $|f_1|$, $|f_2|$ and $|g|$, implies, using (9.11), that $\mathbb{E}\,|Z(t)| < \infty$ and that $\widetilde{Z}(t)$ is a uniformly integrable martingale on any finite interval $[0, T]$.

From (9.4), (9.13), (2.13) and Lemma 9.3 follows

$$\begin{aligned}\mathbb{E}\,\mathcal{Y}(t)\mathcal{Y}'(t) &= \mathbb{E}[\mathcal{Y},\mathcal{Y}]_t + \mathbb{E}\,\mathcal{Y}(0)\mathcal{Y}'(0)\\ &= \sum_{i=1}^q \int_0^t e^{-sA} B_i e^{-sA'} a_i\,\mathbb{E}\,\mathcal{X}_i(s)\,ds + \mathbb{E}\,\mathcal{Y}(0)\mathcal{Y}'(0).\end{aligned} \tag{9.17}$$

By (9.9), this yields the following formula from [9, §V.7.3].

$$\mathbb{E}\,\mathcal{X}(t)\mathcal{X}'(t) = \sum_{i=1}^q \int_0^t e^{(t-s)A} B_i e^{(t-s)A'} a_i\,\mathbb{E}\,\mathcal{X}_i(s)\,ds + e^{tA}\,\mathbb{E}\,\mathcal{X}(0)\mathcal{X}'(0)e^{tA'}. \tag{9.18}$$

We apply the generalized eigenspace decomposition given by $\{P_\lambda\}$ to (9.17) and obtain the following estimate. (See also Lemma 10.6 below.)

Lemma 9.4. (i) *If* $\operatorname{Re}\lambda \le \lambda_1/2$, *then*

$$\mathbb{E}\,|P_\lambda\mathcal{Y}(t)|^2 = O\big((1+t)^{2d_\lambda+1} e^{(\lambda_1 - 2\operatorname{Re}\lambda)t}\big).$$

(ii) *If* $\operatorname{Re}\lambda > \lambda_1/2$, *then*

$$\mathbb{E}\,|P_\lambda\mathcal{Y}(t)|^2 = O(1).$$

Proof. By (9.17), (9.11) and (2.10),

$$\begin{aligned}\big|\mathbb{E}\big(P_\lambda\mathcal{Y}(t)\mathcal{Y}'(t)P'_\lambda\big)\big| &\le C_1\int_0^t \big|P_\lambda e^{-sA}\big|\,e^{\lambda_1 s}\,\big|e^{-sA'}P'_\lambda\big|\,ds + C_2\\ &\le C_3\int_0^t (1+s)^{2d_\lambda} e^{(\lambda_1-2\operatorname{Re}\lambda)s}\,ds + C_2.\end{aligned}$$

The result follows by integration. □

An immediate consequence of Lemmas 9.2 and 9.4(ii) and the martingale convergence theorem for L^2-bounded martingales is the following [9, Theorem V.8.2]. We let $\mathcal{F}_t$ denote the σ-field generated by $\mathcal{X}(s)$, $0 \le s \le t$.

Lemma 9.5. *If* $\operatorname{Re}\lambda > \lambda_1/2$, *then there exists a random vector* $\widetilde{W}_\lambda \in E_\lambda$ *such that* $P_\lambda\mathcal{Y}(t) \to \widetilde{W}_\lambda$ *as* $t\to\infty$, *a.s. and in* L^2. *Moreover,* $P_\lambda\mathcal{Y}(t) = \mathbb{E}(\widetilde{W}_\lambda \mid \mathcal{F}_t)$. □

We can now prove part (iii) of Theorem 3.1. By (9.9) and (2.8),

$$e^{-t\lambda}Q_{\lambda,k}P_\lambda\mathcal{X}(t) = e^{-t\lambda}Q_{\lambda,k}P_\lambda e^{tA}\mathcal{Y}(t) = \sum_{j=0}^{k}\frac{t^j}{j!}Q_{\lambda,k}N_\lambda^j P_\lambda\mathcal{Y}(t).$$

The first part of (iii) of Theorem 3.1 now follows from Lemma 9.5, with $W_{\lambda,k} := (k!)^{-1}Q_{\lambda,k}N_\lambda^k\widetilde{W}_\lambda$. The second part (convergence in distribution to a constant process) follows because the first part implies uniform convergence a.s. for $a \le x \le b$, on every compact interval $[a,b] \subset (0,\infty)$.

Before proving parts (i) and (ii), we observe that the limit in (iii) is non-trivial.

Lemma 9.6. *If $\lambda \in \Lambda_{III}$ and $0 \le k \le d_\lambda$, then $W_{\lambda,k}$ is non-degenerate. More precisely, $\mathbb{P}(W_{\lambda,k} = 0) < 1$ and $\mathbb{P}(W_{\lambda,k} = w) = 0$ for every $w \neq 0$.*

Proof. If $W_{\lambda,k} = 0$ a.s., then $N_\lambda^k\widetilde{W}_\lambda \in N_\lambda^{k+1}E_\lambda$ and thus $N_\lambda^{d_\lambda}\widetilde{W}_\lambda = 0$ a.s.; hence, by Lemma 9.5, for any t,

$$N_\lambda^{d_\lambda}P_\lambda\mathcal{Y}(t) = \mathbb{E}(N_\lambda^{d_\lambda}\widetilde{W}_\lambda \mid \mathcal{F}_t) = 0 \quad \text{a.s.}$$

and thus

$$N_\lambda^{d_\lambda}P_\lambda\mathcal{X}(t) = e^{tA}N_\lambda^{d_\lambda}P_\lambda\mathcal{Y}(t) = 0 \quad \text{a.s.} \tag{9.19}$$

Considering first rational t and using the right-continuity, we see that (9.19) a.s. holds for all $t \ge 0$. It follows that a.s. $N_\lambda^{d_\lambda}P_\lambda\Delta\mathcal{X}(t) = 0$ for all $t > 0$, and thus $N_\lambda^{d_\lambda}P_\lambda\xi_i = 0$ a.s. for every i with $a_i > 0$, since every such transition occurs with positive probability. Taking the expectation we find, since $a_i\,\mathbb{E}\,\xi_i$ equals the i:th column of A by definition, $N_\lambda^{d_\lambda}P_\lambda A = 0$. By (2.5) this yields

$$0 = N_\lambda^{d_\lambda}P_\lambda A = \lambda N_\lambda^{d_\lambda}P_\lambda.$$

Since $N_\lambda^{d_\lambda}P_\lambda \neq 0$ (it is P_λ when $d_\lambda = 0$ and $N_\lambda^{d_\lambda}$ otherwise), this implies $\lambda = 0$, contradicting $\lambda \in \Lambda_{III}$. Consequently, $\mathbb{P}(W_{\lambda,k} = 0) < 1$.

Next, let τ be the time of the first death, and let $\widehat{\mathcal{X}}(t) := \mathcal{X}(\tau + t)$. Then $\widehat{\mathcal{X}}$ is a branching process with the same transitions as $\mathcal{X}$ but a different (random) initial state $\widehat{\mathcal{X}}(0) = \mathcal{X}(\tau)$; moreover, $\widehat{\mathcal{X}}$ is independent of τ. Letting $\widehat{W}_{\lambda,k}$ denote the limit corresponding to $W_{\lambda,k}$ for $\widehat{\mathcal{X}}$, one easily finds $W_{\lambda,k} = e^{-\lambda\tau}\widehat{W}_{\lambda,k}$ a.s. Since τ and $\widehat{W}_{\lambda,k}$ are independent and τ has a continuous distribution, $\mathbb{P}(W_{\lambda,k} = w) = 0$ follows for $w \neq 0$ by conditioning on $\widehat{W}_{\lambda,k}$. □

Because $\lambda_1 > 0$ by our assumption (A3), Lemma 9.5 applies in particular to $\lambda = \lambda_1$. Note that $W_{\lambda_1} = W_{\lambda_1,0} = \widetilde{W}_{\lambda_1}$. We write $W := u_1 \cdot W_{\lambda_1} = u_1' W_{\lambda_1}$ and have by (2.7)

$$W_{\lambda_1} = P_{\lambda_1}W_{\lambda_1} = v_1 u_1' W_{\lambda_1} = v_1 W = W v_1, \tag{9.20}$$

as asserted in Theorem 3.1.

Some well-known properties of W are collected in the next lemma, see [9, Theorems V.6.2 and V.7.2].

Lemma 9.7. *We have $W \geq 0$ a.s., $\mathbb{P}(W > 0) > 0$ and $\mathbb{P}(W = w) = 0$ for every $w \neq 0$. Moreover,*

(i) *As $t \to \infty$, $u_1 \cdot \mathcal{Y}(t) = e^{-\lambda_1 t} u_1 \cdot \mathcal{X}(t) \overset{\text{a.s.}}{\to} W$.*
(ii) *If $u_1 \cdot \mathcal{Y}(t) = 0$ for some t then a.s. $W = 0$.*
(iii) *Conversely, $\mathbb{P}(u_1 \cdot \mathcal{Y}(t) > 0$ and $W = 0) \to 0$ as $t \to \infty$.*

Consequently, a.s. $W = 0$ if and only if $\mathcal{X}$ becomes essentially extinct.

Proof. Since $P'_{\lambda_1} u_1 = u_1$, we have by Lemma 9.5 and (9.20), (2.3)

$$u_1 \cdot \mathcal{Y}(t) = P'_{\lambda_1} u_1 \cdot \mathcal{Y}(t) = u_1 \cdot P_{\lambda_1} \mathcal{Y}(t) \overset{\text{a.s.}}{\to} u_1 \cdot W_{\lambda_1} = u_1 \cdot v_1 W = W.$$

Moreover, $A' u_1 = \lambda_1 u_1$ and thus

$$u_1 \cdot \mathcal{Y}(t) = u_1 \cdot e^{-tA} \mathcal{X}(t) = e^{-tA'} u_1 \cdot \mathcal{X}(t) = e^{-\lambda_1 t} u_1 \cdot \mathcal{X}(t),$$

proving (i). Since $u_1 \cdot \mathcal{X}(t) \geq 0$, $W \geq 0$ follows. If $W = 0$ a.s., then $W_{\lambda_1} = 0$ a.s. by (9.20), which contradicts Lemma 9.6. Similarly, if $w \neq 0$ and $\mathbb{P}(W = w) > 0$, then $\mathbb{P}(W_{\lambda_1} = w v_1) > 0$, which again contradicts Lemma 9.6.

For the second part, recall that $u_{1i} > 0$ for $i \in \mathcal{C}_1$ and $u_{1i} = 0$ for $i \notin \mathcal{C}_1$. Hence, $u_1 \cdot \mathcal{Y}(t) = e^{-\lambda_1 t} u_1 \cdot \mathcal{X}(t) = 0$ if and only if $\mathcal{X}_i(t) = 0$ for all $i \in \mathcal{C}_1$. (In other words, $\mathcal{X}$ is essentially extinct by time t.) Letting $\widetilde{\mathcal{X}}$ denote the branching process obtained from $\mathcal{X}$ by ignoring all balls of types not in $\mathcal{C}_1$, this means that $\widetilde{\mathcal{X}}(t) = 0$. As a consequence, for all $x \geq t$, $\widetilde{\mathcal{X}}(x) = 0$ and thus $u_1 \cdot \mathcal{Y}(x) = 0$; hence the limit $W = 0$.

For the converse, we again consider the process $\widetilde{\mathcal{X}}(t)$. This is an irreducible continuous time branching process, and by [9, Theorem V.7.2], $W = 0$ a.s. implies extinction, $\widetilde{\mathcal{X}}(t) = 0$ for large t, and thus $u_1 \cdot \mathcal{Y}(t) = 0$ for large t. The result follows. □

We may further improve this result to the first claim in Theorem 3.1.

Lemma 9.8. *As $t \to \infty$, $e^{-\lambda_1 t} \mathcal{X}(t) \overset{\text{a.s.}}{\to} W v_1$.*

Proof. This is Theorem V.7.2 in [9], see also [4], but since our setting is somewhat more general, we give a complete proof.

Fix an eigenvalue $\lambda \neq \lambda_1$ and let $\delta := \lambda_1 - \operatorname{Re} \lambda > 0$. Let $\varepsilon > 0$ and let $\mathcal{E}_n$ be the event $\sup_{t \in [n-1,n]} \bigl| e^{-\lambda_1 t} P_\lambda \mathcal{X}(t) \bigr| > \varepsilon$.

If $\mathcal{E}_n$ occurs, then, by (2.10), for some $t \in [n-1, n]$,

$$\varepsilon e^{\lambda_1 t} < |P_\lambda \mathcal{X}(t)| = \bigl| P_\lambda e^{tA} P_\lambda \mathcal{Y}(t) \bigr| \leq C n^{d_\lambda} e^{t(\lambda_1 - \delta)} |P_\lambda \mathcal{Y}(t)|$$

and thus $|P_\lambda \mathcal{Y}(t)| \geq c \varepsilon n^{-d_\lambda} e^{t\delta}$. Consequently, using Doob's inequality and Lemma 9.4,

$$\begin{aligned} \mathbb{P}(\mathcal{E}_n) &\leq \mathbb{P}\bigl(\sup_{t \leq n} |P_\lambda \mathcal{Y}(t)| \geq c \varepsilon n^{-d_\lambda} e^{(n-1)\delta} \bigr) \\ &\leq C \varepsilon^{-2} n^{2d_\lambda} e^{-2n\delta} \, \mathbb{E} \, |P_\lambda \mathcal{Y}(n)|^2 \\ &\leq C \varepsilon^{-2} n^{4d_\lambda + 1} e^{-2n\delta} e^{\max(0, 2\delta - \lambda_1) n}. \end{aligned}$$

Hence $\sum_{n=1}^{\infty} \mathbb{P}(\mathcal{E}_n) < \infty$, so by the Borel–Cantelli lemma a.s. $\limsup_{t\to\infty} |e^{-\lambda_1 t} P_\lambda \mathcal{X}(t)| \le \varepsilon$. In other words,

$$e^{-\lambda_1 t} P_\lambda \mathcal{X}(t) \to 0 \text{ a.s.}, \qquad \lambda \neq \lambda_1. \tag{9.21}$$

Moreover, $P_{\lambda_1} e^{-tA} = e^{-\lambda_1 t} P_{\lambda_1}$ and thus, by Lemma 9.5 and (9.20),

$$e^{-\lambda_1 t} P_{\lambda_1} \mathcal{X}(t) = P_{\lambda_1} \mathcal{Y}(t) \overset{\text{a.s.}}{\to} \widetilde{W}_{\lambda_1} = W_{\lambda_1} = W v_1. \tag{9.22}$$

The result follows by (9.21), (9.22) and (2.4). □

Lemma 9.9. *Let, for $t, y \ge 0$, $Z_y(t)$ be defined as in Lemma 9.3 using matrix functions f_{1y}, f_{2y} and g. Suppose further that, as $t \to \infty$ and for any fixed T,*

$$\sup_{t>1} \int_0^t |f_{1t}(s)|\,|f_{2t}(s)| e^{\lambda_1 s}\, ds < \infty, \tag{9.23}$$

$$\int_0^t |f_{1t}(s)|^2 |f_{2t}(s)|^2 e^{\lambda_1 s}\, ds \to 0, \tag{9.24}$$

$$\int_0^T |f_{1t}(s)|\,|f_{2t}(s)|\, ds \to 0. \tag{9.25}$$

Then, as $t \to \infty$,

$$Z_t(t) - a_i v_{1i} W \int_0^t f_{1t}(s)\, \mathbb{E}\, g(\xi_i) f_{2t}(s) e^{\lambda_1 s}\, ds \overset{\text{p}}{\to} 0.$$

Proof. We begin by showing that, defining $\widetilde{Z}_t$ as in (9.15),

$$\mathbb{E}\, |\widetilde{Z}_t(t)| \to 0. \tag{9.26}$$

By considering the components separately, and taking real and imaginary parts, we may assume that f_{1y}, f_{2y}, g and Z_y are real-valued; we write $f_y = f_{1y} f_{2y}$.

Assume first that $\mathbb{E}\, |g(\xi_i)|^2 < \infty$. Since the martingale $\widetilde{Z}_y(t)$, $t \ge 0$, has finite variation, its quadratic variation is by (9.1) given by

$$[\widetilde{Z}_y, \widetilde{Z}_y]_t = \sum_{s \le t} \bigl(\Delta Z_y(t)\bigr)^2 = \sum_{k:\tau_{ik} \le t} f_y(\tau_{ik})^2 g\bigl(\Delta \mathcal{X}(\tau_{ik})\bigr)^2.$$

This is a sum of the same type as Z, with f_1, g and f_2 replaced by f_y^2, g^2 and 1. Hence Lemma 9.3 yields, together with (9.3) and (9.11),

$$\begin{aligned}
\mathbb{E}\, \widetilde{Z}_y(t)^2 = \mathbb{E}[\widetilde{Z}_y, \widetilde{Z}_y]_t &= \int_0^t f_y(s)^2\, \mathbb{E}\, g(\xi_i)^2 a_i\, \mathbb{E}\, \mathcal{X}_i(s)\, ds \\
&\le C \int_0^t f_y(s)^2 e^{\lambda_1 s}\, ds = C \int_0^t f_{1y}(s)^2 f_{2y}(s)^2 e^{\lambda_1 s}\, ds
\end{aligned}$$

and thus, by (9.24), $\mathbb{E}\, |\widetilde{Z}_t(t)|^2 \to 0$, which proves (9.26).

In the case $\mathbb{E}\, |g(\xi_i)|^2 = \infty$ we truncate, defining $g_1(x) = g(x)\mathbf{1}[|g(x)| \le M]$ and $g_2(x) = g(x) - g_1(x)$ for a constant M. Thus $g = g_1 + g_2$ and there is a corresponding decomposition $\widetilde{Z}_y(t) = \widetilde{Z}_{1y}(t) + \widetilde{Z}_{2y}(t)$, where $\mathbb{E}\,|\widetilde{Z}_{1t}(t)| \to 0$ by the case just proved. Applying Lemma 9.3 with $|f_{1y}|$, $|f_{2y}|$ and $|g_2|$,

$$\begin{aligned}\mathbb{E}\,|\widetilde{Z}_{2y}(t)| &\le \mathbb{E}\,|Z_{2y}(t)| + \int_0^t |f_{1y}(s)|\,\mathbb{E}\,|g_2(\xi_i)||f_{2y}(s)|a_i\,\mathbb{E}\,\mathcal{X}_i(s)\,ds \\ &\le 2\int_0^t |f_{1y}(s)|\,\mathbb{E}\,|g_2(\xi_i)||f_{2y}(s)|a_i\,\mathbb{E}\,\mathcal{X}_i(s)\,ds.\end{aligned}$$

Hence, by (9.11) and (9.23), for $t \ge 1$,

$$\mathbb{E}\,|\widetilde{Z}_{2t}(t)| \le C_1\,\mathbb{E}\,|g_2(\xi_i)| \int_0^t |f_{1t}(s)|\,|f_{2t}(s)|e^{\lambda_1 s}\,ds \le C_2\,\mathbb{E}\,|g_2(\xi_i)|$$

and thus

$$\limsup_{t\to\infty} \mathbb{E}\,|\widetilde{Z}_t(t)| = \limsup_{t\to\infty} \mathbb{E}\,|\widetilde{Z}_{2t}(t)| \le C_2\,\mathbb{E}\,|g_2(\xi_i)|.$$

Since $\mathbb{E}\,|g_2(\xi_i)|$ can be made arbitrarily small by choosing M large, (9.26) follows.

Next, for every $\varepsilon > 0$, there is by Lemma 9.8 a.s. a random T such that $|e^{-\lambda_1 t}\mathcal{X}_i(t) - v_{1i}W| < \varepsilon$ for $t \ge T$. Hence, for $t \ge T$, for some random K independent of t and again using (9.23),

$$\begin{aligned}&\left|\int_0^t f_{1t}(s)\,\mathbb{E}\,g(\xi_i)f_{2t}(s)a_i\big(\mathcal{X}_i(s) - v_{1i}We^{\lambda_1 s}\big)ds\right| \\ &\le C_3\int_0^T |f_{1t}(s)||f_{2t}(s)|\big(\mathcal{X}_i(s) + v_{1i}We^{\lambda_1 s}\big)ds + C_4\int_T^t |f_{1t}(s)||f_{2t}(s)|\varepsilon e^{\lambda_1 s}\,ds \\ &\le K\int_0^T |f_{1t}(s)||f_{2t}(s)|\,ds + C_5\varepsilon.\end{aligned}$$

It follows by (9.25) and the arbitrariness of ε that

$$\int_0^t f_{1t}(s)\,\mathbb{E}\,g(\xi_i)f_{2t}(s)a_i\big(\mathcal{X}_i(s) - v_{1i}We^{\lambda_1 s}\big)ds \to 0 \text{ a.s.,} \tag{9.27}$$

and thus in probability, as $t \to \infty$. The result follows from (9.15), (9.26) and (9.27). □

Lemma 9.9 and (9.13) will give us the estimates of $[\mathcal{Y},\mathcal{Y}]_t$ that we need in order to apply Proposition 9.1. We begin with $P_I\mathcal{Y}$, the part corresponding to eigenvalues λ with $\operatorname{Re}\lambda < \lambda_1/2$. For convenience we set $\widetilde{A} := P_I A - \lambda_1/2$, and note that each eigenvalue μ of $\widetilde{A}$ satisfies $\operatorname{Re}\mu < 0$. Hence, for some $\delta > 0$ and every $t \ge 0$,

$$|e^{t\widetilde{A}}| = O(e^{-\delta t}). \tag{9.28}$$

Define, recalling (9.9),

$$\mathcal{Y}_y(t) := e^{y\widetilde{A}}P_I\mathcal{Y}(t) = e^{-\lambda_1 y/2 + (y-t)A}P_I\mathcal{X}(t). \tag{9.29}$$

For fixed y, $\mathcal{Y}_y(t)$ is a martingale by Lemma 9.2, and, by (9.13),

$$[\mathcal{Y}_y,\mathcal{Y}_y]_t = e^{y\widetilde{A}}P_I[\mathcal{Y},\mathcal{Y}]_tP_I'e^{y\widetilde{A}'} = \sum_{i=1}^q \sum_{k:\tau_{ik}\le t} f_{1y}^{(i)}(\tau_{ik})g\bigl(\Delta\mathcal{X}(\tau_{ik})\bigr)f_{2y}^{(i)}(\tau_{ik}), \tag{9.30}$$

where $g(\xi)=\xi\xi'$ and

$$f_{1y}^{(i)}(s) = f_{2y}^{(i)\prime}(s) = e^{y\widetilde{A}}P_Ie^{-sA} = e^{-\lambda_1 s/2+(y-s)\widetilde{A}}P_I. \tag{9.31}$$

The inner sum in (9.30) is of the type studied in Lemmas 9.3 and 9.9. We apply Lemma 9.9 for each i separately. By (9.31) and (9.28),

$$|f_{1t}^{(i)}(s)| = |f_{2t}^{(i)}(s)| = O\bigl(e^{-\lambda_1 s/2-(t-s)\delta}\bigr), \qquad 0\le s\le t, \tag{9.32}$$

and the conditions (9.23)–(9.25) follow.

Moreover, $\mathbb{E}\,g(\xi_i)=B_i$ by (2.13), and by (9.31)

$$\begin{aligned}\int_0^t f_{1t}^{(i)}(s)\,\mathbb{E}\,g(\xi_i)f_{2t}^{(i)}(s)e^{\lambda_1 s}\,ds &= \int_0^t e^{(t-s)\widetilde{A}}P_IB_iP_I'e^{(t-s)\widetilde{A}'}\,ds\\ &= P_I\int_0^t e^{r\widetilde{A}}B_ie^{r\widetilde{A}'}\,dr\,P_I' \to P_I\int_0^\infty e^{r\widetilde{A}}B_ie^{r\widetilde{A}'}\,dr\,P_I',\end{aligned}$$

where the integral converges by (9.28). Consequently, (9.30) and Lemma 9.9 yield, as $t\to\infty$, see (2.15),

$$[\mathcal{Y}_t,\mathcal{Y}_t]_t \overset{\mathrm{p}}{\to} \sum_{i=1}^q a_iv_{1i}WP_I\int_0^\infty e^{s\widetilde{A}}B_ie^{s\widetilde{A}'}\,ds\,P_I' = W\Sigma_I. \tag{9.33}$$

For any fixed real x, $\mathcal{Y}_t(s)=e^{-x\widetilde{A}}\mathcal{Y}_{t+x}(s)$ and thus, as $t\to\infty$,

$$[\mathcal{Y}_t,\mathcal{Y}_t]_{t+x} = e^{-x\widetilde{A}}[\mathcal{Y}_{t+x},\mathcal{Y}_{t+x}]_{t+x}e^{-x\widetilde{A}'} \overset{\mathrm{p}}{\to} We^{-x\widetilde{A}}\Sigma_Ie^{-x\widetilde{A}'}. \tag{9.34}$$

Moreover, by (9.30) and Lemma 9.3

$$\mathbb{E}[\mathcal{Y}_t,\mathcal{Y}_t]_t = \sum_{i=1}^q\int_0^t f_{1t}^{(i)}(s)B_if_{2t}^{(i)}(s)a_i\,\mathbb{E}\,\mathcal{X}_i(s)\,ds$$

which by (9.32) and (9.11) is bounded for $t\ge0$. Consequently, for $t+x\ge0$,

$$\mathbb{E}[\mathcal{Y}_t,\mathcal{Y}_t]_{t+x} = e^{-x\widetilde{A}}\,\mathbb{E}[\mathcal{Y}_{t+x},\mathcal{Y}_{t+x}]_{t+x}e^{-x\widetilde{A}'} = O\bigl(|e^{-x\widetilde{A}}|^2\bigr).$$

We have further

$$|\mathcal{Y}_t(0)| = |e^{t\widetilde{A}}P_I\mathcal{Y}(0)| = O(|e^{t\widetilde{A}}|).$$

and thus, using (9.3) on each component of $\mathcal{Y}_t$,

$$\mathbb{E}\,|\mathcal{Y}_t(t+x)|^2 = |\mathcal{Y}_t(0)|^2+\operatorname{Tr}\mathbb{E}[\mathcal{Y}_t,\mathcal{Y}_t]_{t+x} = O\bigl(|e^{t\widetilde{A}}|^2+|e^{-x\widetilde{A}}|^2\bigr). \tag{9.35}$$

In particular, for any fixed x, by (9.28),

$$\sup_{t\ge-x}\mathbb{E}\,|\mathcal{Y}_t(t+x)|^2<\infty. \tag{9.36}$$

We are close to applying Proposition 9.1, but we still have the obstacle that the right-hand side of (9.34) is random. To simply divide $\mathcal{Y}_t$ by $\sqrt{W}$ would destroy the martingale property, but we can do something similar: Let, for $t \geq 1$, $h = h(t) := t^{1/2}$ (any function of t that increases slowly to ∞ would do), and define (with the convention $0^{-1/2} = 0$)

$$\widetilde{Y}_t(x) := \begin{cases} \bigl(u_1 \cdot \mathcal{Y}(t-h)\bigr)^{-1/2}\bigl(\mathcal{Y}_t(t+\ln x) - \mathcal{Y}_t(t-h)\bigr), & x \geq e^{-h}, \\ 0, & x < e^{-h}. \end{cases} \tag{9.37}$$

Clearly, $\widetilde{Y}_t$ is a martingale on $[0,\infty)$, and this is still true conditioned on any event $\mathcal{E}_t \in \mathcal{F}_{t-h}$.

For fixed $x > 0$ we have from (9.1), for t so large that $-h(t) < \ln x$,

$$[\widetilde{Y}_t, \widetilde{Y}_t]_x = \bigl(u_1 \cdot \mathcal{Y}(t-h)\bigr)^{-1}\bigl([\mathcal{Y}_t, \mathcal{Y}_t]_{t+\ln x} - [\mathcal{Y}_t, \mathcal{Y}_t]_{t-h}\bigr). \tag{9.38}$$

Further, using Doob's inequality, (9.35) and (9.28),

$$\mathbb{E} \sup_{x \leq t-h} |\mathcal{Y}_t(x)|^2 \leq 4\,\mathbb{E}\,|\mathcal{Y}_t(t-h)|^2 = O\bigl(e^{-2\delta t} + e^{-2\delta h}\bigr) = o(1), \tag{9.39}$$

as $t \to \infty$. It follows by (9.3) applied to the components $\mathcal{Y}_{ti}$ that, for $i = 1,\dots,q$, $\mathbb{E}[\mathcal{Y}_{ti}, \mathcal{Y}_{ti}]_{t-h} \to 0$. The same holds for the nondiagonal entries of $\mathbb{E}[\mathcal{Y}_t, \mathcal{Y}_t]_{t-h}$ by the Cauchy–Schwarz inequality and the Kunita–Watanabe inequality $|[\mathcal{Y}_{ti}, \mathcal{Y}_{tj}]_s| \leq [\mathcal{Y}_{ti}, \mathcal{Y}_{ti}]_s^{1/2}[\mathcal{Y}_{tj}, \mathcal{Y}_{tj}]_s^{1/2}$, which in this setting is the Cauchy–Schwarz inequality applied to (9.1). Hence,

$$\mathbb{E}[\mathcal{Y}_t, \mathcal{Y}_t]_{t-h} \to 0. \tag{9.40}$$

Since $t - h \to \infty$, Lemma 9.7 yields $u_1 \cdot \mathcal{Y}(t-h) \overset{\text{a.s.}}{\to} W$. Combined with (9.38), (9.34) and (9.40), this implies that, on the event $\{W > 0\}$, for every fixed $x > 0$,

$$[\widetilde{Y}_t, \widetilde{Y}_t]_x \overset{\text{p}}{\to} e^{-(\ln x)\widetilde{A}}\Sigma_I e^{-(\ln x)\widetilde{A}'}. \tag{9.41}$$

Now assume that $\varepsilon > 0$ and that, for large t, $\mathcal{E}_t$ is any event in $\mathcal{F}_{t-h}$ with $\mathbb{P}(\mathcal{E}_t) \geq \varepsilon$ and $u_1 \cdot \mathcal{Y}(t-h) \geq \varepsilon$ on $\mathcal{E}_t$. Since $\mathbb{P}(\mathcal{E}_t \cap \{W = 0\}) \to 0$ by Lemma 9.7, it follows from (9.41), (9.36) and (9.39) that $\widetilde{Y}_t$ conditioned on $\mathcal{E}_t$ satisfies the conditions of Proposition 9.1 with $\Sigma(x) := e^{-(\ln x)\widetilde{A}}\Sigma_I e^{-(\ln x)\widetilde{A}'}$. ($\Sigma(0) = 0$; note that Σ is continuous at 0 by (9.28).) Consequently, conditioned on $\mathcal{E}_t$, $\widetilde{Y}_t(x) \overset{\text{d}}{\to} \widetilde{U}_I(x)$ in $D[0,\infty)$ as $t \to \infty$, where $\widetilde{U}_I$ is the continuous vector-valued Gaussian stochastic process with mean 0 and covariances

$$\mathbb{E}\,\widetilde{U}_I(x)\widetilde{U}_I(y)' = \mathbb{E}\,\widetilde{U}_I(x)\widetilde{U}_I(x)' = \Sigma(x) = e^{-(\ln x)\widetilde{A}}\Sigma_I e^{-(\ln x)\widetilde{A}'}, \qquad x \leq y. \tag{9.42}$$

Next, from the definition of $\widetilde{Y}_t$ and (9.39), on $\mathcal{E}_t$,

$$\sup_{x \geq 0}\bigl|\widetilde{Y}_t(x) - \bigl(u_1 \cdot \mathcal{Y}(t-h)\bigr)^{-1/2}\mathcal{Y}_t(t+\ln x)\bigr| = \bigl(u_1 \cdot \mathcal{Y}(t-h)\bigr)^{-1/2} \sup_{y \leq t-h} |\mathcal{Y}_t(y)| \overset{\text{p}}{\to} 0$$

so we may replace $\widetilde{Y}_t(x)$ by $\bigl(u_1 \cdot \mathcal{Y}(t-h)\bigr)^{-1/2}\mathcal{Y}_t(t+\ln x)$ in this limit result. Changing variables $x \to e^x$ we find, in $D[-\infty,\infty)$ and thus in $D(-\infty,\infty)$, conditioned on $\mathcal{E}_t$, as $t \to \infty$,

$$\big(u_1 \cdot \mathcal{Y}(t-h)\big)^{-1/2} \mathcal{Y}_t(t+x) \xrightarrow{\mathrm{d}} \widetilde{U}_I(e^x).$$

Since, by (9.29), $\mathcal{Y}_t(t+x) = e^{-\lambda_1 t/2 - xA} P_I \mathcal{X}(t+x)$, we obtain by multiplication by $e^{x\widetilde{A}}$, as $t \to \infty$ and conditioned on $\mathcal{E}_t$, in $D(-\infty,\infty)$,

$$\big(u_1 \cdot \mathcal{Y}(t-h)\big)^{-1/2} e^{-\lambda_1(t+x)/2} P_I \mathcal{X}(t+x) \xrightarrow{\mathrm{d}} U_I(x) := e^{x\widetilde{A}} \widetilde{U}_I(e^x).$$

U_I is a Gaussian process, and the covariances (3.1) follow from (9.42).

We now apply Lemma 8.2 taking $\eta_t = u_1 \cdot \mathcal{Y}(t-h)$, $\zeta_t = \big(u_1 \cdot \mathcal{Y}(t-h)\big)^{-1/2} e^{-\lambda_1(t+x)/2} P_I \mathcal{X}(t+x)$, $\eta = W$, $\zeta = U_I$, $\mathcal{S}_1 = \mathbb{R}$, $\mathcal{S}_2 = D(-\infty,\infty)$, $E = \{r \in \mathbb{R} : r > 0\}$, and $E_m = \{r \in \mathbb{R} : r > 1/m\}$.

We have just proved assumption (iv) in Lemma 8.2, and assumptions (i), (ii) and (iii) follow by Lemma 9.7. Hence Lemma 8.2 shows that, conditioned on $u_1 \cdot \mathcal{Y}(t-h) > 0$, $(\eta_t, \zeta_t) \xrightarrow{\mathrm{d}} (\widetilde{W}, U_I)$, where $\widetilde{W}$ has the distribution of W conditioned on being positive and with U_I independent of $\widetilde{W}$. Consequently, conditioned on $u_1 \cdot \mathcal{Y}(t-h) > 0$,

$$e^{-\lambda_1(t+x)/2} P_I \mathcal{X}(t+x) = \eta_t^{1/2} \zeta_t \xrightarrow{\mathrm{d}} \widetilde{W}^{1/2} U_I(x) \quad \text{in } D(-\infty,\infty). \tag{9.43}$$

If $\mathbb{P}(W = 0) = 0$, then $\widetilde{W} = W$ and $u_1 \cdot \mathcal{Y}(t-h) > 0$ a.s. by Lemma 9.7, and part (i) of Theorem 3.1 is proved.

If $\mathbb{P}(W = 0) > 0$, we define the martingale

$$\widehat{Y}_t(x) := \begin{cases} \mathcal{Y}_t(t + \ln x) - \mathcal{Y}_t(t-h), & x \ge e^{-h}, \\ 0, & x < e^{-h}. \end{cases}$$

Since $u_1 \cdot \mathcal{Y}(t-h) = 0$ implies $W = 0$ a.s. by Lemma 9.7, (9.34) implies that, conditioned on $u_1 \cdot \mathcal{Y}(t-h) = 0$,

$$[\widehat{Y}_t, \widehat{Y}_t]_x \xrightarrow{\mathrm{p}} 0$$

for every fixed x. Together with (9.36) and (9.39), this shows that we can apply Proposition 9.1 to $\widehat{Y}_t(x)$ conditioned on $u_1 \cdot \mathcal{Y}(t-h) = 0$, now with $\Sigma(x) = 0$ and thus $M = 0$. Hence, conditioned on $u_1 \cdot \mathcal{Y}(t-h) = 0$, we find first $\widehat{Y}_t(x) \xrightarrow{\mathrm{d}} 0$ in $D[0,\infty)$, then by (9.39) and a change of variable $\mathcal{Y}_t(t+x) \xrightarrow{\mathrm{d}} 0$ in $D(-\infty,\infty)$, and then, after multiplication by $e^{x\widetilde{A}'}$,

$$e^{-\lambda_1(t+x)/2} P_I \mathcal{X}(t+x) \xrightarrow{\mathrm{d}} 0 = W^{1/2} U_I(x) \quad \text{in } D(-\infty,\infty).$$

This complements (9.43) and together they imply part (i) of Theorem 3.1, with U_I independent of W.

A simple modification of the argument shows that (i) and (iii) hold jointly. (When we talk about (iii) holding jointly with other assertions, here and later in the proof, we only mean the part on convergence in distribution.) Indeed, let $\mathcal{Y}_{III,t}(xt)$ denote the collection of the left-hand sides of (iii) in Theorem 3.1 (for $\lambda \in \Lambda_{III}$ and $0 \le k \le d_\lambda$), regarded as a large vector-valued process, and let W_{III} denote the corresponding collection of the right-hand sides. The a.s. convergence in (iii) implies that $\mathcal{Y}_{III,t}(xt) - \mathcal{Y}_{III,t}(h) \overset{\mathrm{p}}{\to} 0$ in $D(0,\infty)$ as functions of x. Hence the convergence in distribution in (iii) is equivalent to $\mathcal{Y}_{III,t}(h) \overset{\mathrm{d}}{\to} W_{III}$.

By enlarging η_t (and $\mathcal{S}_1$) in the application of Lemma 8.2 above, changing η_t into $\bigl(u_1 \cdot \mathcal{Y}(t-h), \mathcal{Y}_{III,t}(h)\bigr)$, it now follows that (i) and (iii) hold jointly, with U_I independent of $\{W_{\lambda,k}\}_{\lambda\in\Lambda_{III},k\le d_\lambda}$.

For $\lambda \in \Lambda_{II}$ we argue similarly, now putting

$$\mathcal{Y}_{y,\lambda,k}(t) = y^{-k-1/2} Q_{\lambda,k} P_\lambda \mathcal{Y}(t). \tag{9.44}$$

Again, this is a martingale (for fixed y, λ, k). If also $\mu \in \Lambda_{II}$ and $0 \le l \le d_\mu$, then $[\mathcal{Y}_{y,\lambda,k}, \mathcal{Y}_{y,\mu,l}]_t$ is by (9.13) given by a sum as in (9.30), where now by (2.8)

$$f_{1y}^{(i)}(s) = y^{-k-1/2} Q_{\lambda,k} P_\lambda e^{-sA} = y^{-k-1/2} e^{-s\lambda} \sum_{j=0}^{k} \frac{(-s)^j}{j!} Q_{\lambda,k} N_\lambda^j P_\lambda$$

and similarly for $f_{2y}^{(i)}(s)$. Instead of (9.32) we have by (2.11), for $t \ge 1$,

$$|f_{1t}^{(i)}(s)| = O\bigl(t^{-k-1/2} e^{-\lambda_1 s/2} (1+s)^k\bigr) = O\bigl(t^{-1/2} e^{-\lambda_1 s/2}\bigr), \qquad 0 \le s \le t,$$

and similarly for $f_{2y}^{(i)}(s)$. In this case too, (9.23)–(9.25) follow, so Lemma 9.9 applies for each i. We now have

$$\begin{aligned} &\int_0^t f_{1t}^{(i)}(s) \,\mathbb{E}\, g(\xi_i) f_{2t}^{(i)}(s) e^{\lambda_1 s}\, ds \\ &= \int_0^t t^{-k-l-1} e^{s(\lambda_1-\lambda-\mu)} \sum_{j=0}^{k}\sum_{m=0}^{l} \frac{(-s)^{j+m}}{j!\,m!} Q_{\lambda,k} N_\lambda^j P_\lambda B_i P'_\mu (N'_\mu)^m Q'_{\mu,l}\, ds. \end{aligned} \tag{9.45}$$

If $\mu = \bar\lambda$, then $\lambda_1 - \lambda - \mu = \lambda_1 - 2\operatorname{Re}\lambda = 0$, and the integral (9.45) equals

$$\begin{aligned} &\sum_{j=0}^{k}\sum_{m=0}^{l} (-1)^{j+m} \frac{t^{j+m+1-k-l-1}}{(j+m+1)\,j!\,m!} Q_{\lambda,k} N_\lambda^j P_\lambda B_i P'_\mu (N'_\mu)^m Q'_{\mu,l} \\ &\qquad \to \frac{(-1)^{k+l}}{(k+l+1)\,k!\,l!} Q_{\lambda,k} N_\lambda^k P_\lambda B_i P'_\mu (N'_\mu)^l Q'_{\mu,l}, \end{aligned}$$

as $t \to \infty$. If $\mu \ne \bar\lambda$, then $\lambda_1 - \lambda - \mu$ is imaginary and nonzero. It is easily seen, by integration by parts, that then $\int_0^t e^{s(\lambda_1-\lambda-\mu)} s^r\, ds = O(t^r)$ for each $r \ge 0$, which implies that the integral (9.45) tends to 0.

Hence, in both cases, Lemma 9.9 and (2.16) imply that, as $t \to \infty$,

$$[\mathcal{Y}_{t,\lambda,k}, \mathcal{Y}_{t,\mu,l}]_t \overset{\mathrm{p}}{\to} W \cdot \frac{(-1)^{k+l}}{(k+l+1)\,k!\,l!} Q_{\lambda,k} N_\lambda^k P_\lambda \Sigma_{II} P'_\mu (N'_\mu)^l Q'_{\mu,l}$$

(with the right-hand side 0 by (2.16) unless $\mu = \bar{\lambda}$). Consequently, for every fixed $x > 0$,

$$\begin{aligned}[\mathcal{Y}_{t,\lambda,k}, \mathcal{Y}_{t,\mu,l}]_{xt} &= x^{k+l+1}[\mathcal{Y}_{xt,\lambda,k}, \mathcal{Y}_{xt,\mu,l}]_{xt} \\ &\overset{\mathrm{p}}{\to} W \cdot x^{k+l+1} \frac{(-1)^{k+l}}{(k+l+1)\,k!\,l!} Q_{\lambda,k} N_\lambda^k P_\lambda \Sigma_{II} P'_\mu (N'_\mu)^l Q'_{\mu,l}.\end{aligned} \tag{9.46}$$

Clearly, this holds for $x = 0$ too. Moreover, it follows easily from (9.44), (9.17) and (2.11) that $\sup_{t\geq 1} \mathbb{E}\,|\mathcal{Y}_{t,\lambda,k}(xt)|^2 < \infty$ for every $x \geq 0$.

Let $\mathcal{Y}_{II,t}(s)$ be the vector obtained by combining all vectors $\mathcal{Y}_{t,\lambda,k}(s)$ for λ and k as in part (ii) of Theorem 3.1; we write this as $\mathcal{Y}_{II,t} = (\mathcal{Y}_{t,\lambda,k})_{\lambda,k}$. We thus see from (9.46) that $[\mathcal{Y}_{II,t}, \mathcal{Y}_{II,t}]_{xt}$ converges. Moreover, since $\overline{\mathcal{Y}_{t,\lambda,k}} = \mathcal{Y}_{t,\bar{\lambda},k}$, $\overline{\mathcal{Y}_{II,t}}$ equals $\mathcal{Y}_{II,t}$ with a certain permutation of the components. Hence $[\mathcal{Y}_{II,t}, \overline{\mathcal{Y}_{II,t}}]_{xt}$ too converges.

We this time define, with $h := t^{1/2}$ as above,

$$\widetilde{Y}_t(x) := \begin{cases} \big(u_1 \cdot \mathcal{Y}(h)\big)^{-1/2}\big(\mathcal{Y}_{II,t}(xt) - \mathcal{Y}_{II,t}(h)\big), & x \geq h/t, \\ 0, & x < h/t. \end{cases}$$

Repeating the argument after (9.37) above (replacing $t - h$ by $t/h = h$ and using Proposition 9.1(ii) if some $\lambda \in \Lambda_{II}$ is non-real), we obtain, in $D[0,\infty)$,

$$\mathcal{Y}_{II,t}(xt) \overset{\mathrm{d}}{\to} W^{1/2} \widetilde{U}_{II}(x) \tag{9.47}$$

or, jointly for all $\lambda \in \Lambda_{II}$ and $0 \leq k \leq d_\lambda$,

$$\mathcal{Y}_{t,\lambda,k}(xt) \overset{\mathrm{d}}{\to} W^{1/2} \widetilde{U}_{\lambda,k}(x), \tag{9.48}$$

where $\widetilde{U}_{II} = (\widetilde{U}_{\lambda,k})_{\lambda,k}$ is Gaussian and, for $0 \leq x \leq y$,

$$\mathbb{E}\big(\widetilde{U}_{\lambda,k}(x)\widetilde{U}_{\mu,l}(y)'\big) = \tilde{c}(k,l,x) Q_{\lambda,k} N_\lambda^k P_\lambda \Sigma_{II} P'_\mu (N'_\mu)^l Q'_{\mu,l}. \tag{9.49}$$

with

$$\tilde{c}(k,l,x) := \frac{(-1)^{k+l}}{(k+l+1)k!\,l!} x^{k+l+1} = \frac{(-1)^{k+l}}{k!\,l!} \int_0^x s^{k+l}\, ds. \tag{9.50}$$

In the case $d_\lambda = 0$ for $\lambda \in \Lambda_{II}$, this completes the proof of (ii). In general, we have by (9.9), (2.8), (2.6) and (9.44)

$$\begin{aligned}t^{-(k+1/2)} e^{-\lambda xt} Q_{\lambda,k} P_\lambda \mathcal{X}(xt) &= t^{-k-1/2} e^{-\lambda xt} Q_{\lambda,k} P_\lambda e^{xtA} \mathcal{Y}(xt) \\ &= t^{-k-1/2} Q_{\lambda,k} e^{xtN_\lambda} P_\lambda \mathcal{Y}(xt) = t^{-k-1/2} \sum_{j=0}^k \frac{(xt)^j}{j!} Q_{\lambda,k} N_\lambda^j P_\lambda \mathcal{Y}(xt) \\ &= \sum_{j=0}^k \frac{x^j}{j!} N_\lambda^j \mathcal{Y}_{t,\lambda,k-j}(xt),\end{aligned}$$

which by (9.48) yields part (ii) of the theorem with

$$U_{\lambda,k}(x) := \sum_{j=0}^{k} \frac{x^j}{j!} N_\lambda^j \widetilde{U}_{\lambda,k-j}(x).$$

By (9.49) we have (3.2), with, using (9.50) and the binomial theorem,

$$\begin{aligned} c(k,l,x,y) &= \sum_{j=0}^{k}\sum_{m=0}^{l} \frac{x^j}{j!}\frac{y^m}{m!}\tilde{c}(k-j,l-m,x) \\ &= \sum_{j=0}^{k}\sum_{m=0}^{l} \frac{x^j}{j!}\frac{y^m}{m!}\frac{1}{(k-j)!\,(l-m)!}\int_0^x (-s)^{k-j+l-m}\,ds \\ &= \frac{1}{k!\,l!}\int_0^x (x-s)^k(y-s)^l\,ds, \end{aligned}$$

which yields (3.3) by a change of variable and the binomial theorem again.

Again, by Lemma 8.2 with an enlarged η_t as above, (ii) and (iii) hold jointly, with $\{U_{\lambda,k}\}_{\lambda\in\Lambda_{II},\,k\le d_\lambda}$ independent of $\{W_{\lambda,k}\}_{\lambda\in\Lambda_{III},k\le d_\lambda}$.

We have proved (i), (ii) and (iii) separately, but only partly proved the joint convergence and the asserted independence.

First, note that the families $\{U_{\lambda,k}\}_{0\le k\le d_\lambda}$ for different $\lambda \in \Lambda_{II}$ with $\operatorname{Im}\lambda \ge 0$ are independent because they are jointly Gaussian and all covariances of real or imaginary parts vanish by (3.2) and (2.16).

Next, consider the stopped processes $\mathcal{Y}^T_{II,t}(s) := \mathcal{Y}_{II,t}(s\wedge T)$ and $\widetilde{U}^T_{II}(x) := \widetilde{U}_{II}(x\wedge T)$. Stopping at $x=1$ we obtain from (9.47)

$$\mathcal{Y}^t_{II,t}(xt) \xrightarrow{\mathrm{d}} W^{1/2}\widetilde{U}^1_{II}(x). \tag{9.51}$$

Moreover, we have shown that this holds jointly with (iii) and with $\widetilde{U}^1_{II}$ independent of $\{W_{\lambda,k}\}_{\lambda\in\Lambda_{III},k\le d_\lambda}$. It is another consequence of (9.47) that

$$\sup_x\bigl|\mathcal{Y}^t_{II,t}(xt) - \mathcal{Y}^{t-h}_{II,t}(xt)\bigr| = \sup_{1-h/t\le x\le 1}|\mathcal{Y}_{II,t}(xt)-\mathcal{Y}_{II,t}(t-h)| \xrightarrow{\mathrm{p}} 0. \tag{9.52}$$

Hence (9.51) is equivalent to

$$\mathcal{Y}^{t-h}_{II,t}(xt) \xrightarrow{\mathrm{d}} W^{1/2}\widetilde{U}^1_{II}(x). \tag{9.53}$$

Next, redo the application above of Lemma 8.2 to (i) once again, now further enlarging η_t (and $\mathcal{S}_1$) to contain also $\mathcal{Y}^{t-h}_{II,t}(xt)$ (which is $\mathcal{F}_{t-h}$-measurable). This shows that (9.53), (iii) and (i) hold jointly, with U_I, $\widetilde{U}^1_{II}$ and $\{W_{\lambda,k}\}_{\lambda\in\Lambda_{III},k\le d_\lambda}$ independent.

Finally, stop the left-hand side of (i) at $x=h$; this does not affect the limit because $h\to\infty$. Consider again the argument for (9.47), but applied to $\mathcal{Y}_{II,t}-\mathcal{Y}^{t+h}_{II,t}$. Applying Lemma 8.2 as above but with η_t containing $\mathcal{Y}^{t-h}_{II,t}$, $\mathcal{Y}_{III,t}(h)$ and the left-hand side of (i) stopped at $x=h$, we find

$$\mathcal{Y}_{II,t}(xt)-\mathcal{Y}^{t+h}_{II,t}(xt) \xrightarrow{\mathrm{d}} W^{1/2}\bigl(\widetilde{U}_{II}(x)-\widetilde{U}^1_{II}(x)\bigr) \quad \text{in } D[0,\infty), \tag{9.54}$$

jointly with (i), (iii) and (9.53), with $\widetilde{U}_{II} - \widetilde{U}^1_{II}$ independent of U_I, $\widetilde{U}^1_{II}$ and $\{W_{\lambda,k}\}_{\lambda\in\Lambda_{III},k\le d_\lambda}$.

By (9.52) and the analogous $\sup_x |\mathcal{Y}^{t+h}_{II,t}(xt) - \mathcal{Y}^{t}_{II,t}(xt)| \overset{\mathrm{p}}{\to} 0$, we can replace $\mathcal{Y}^{t\pm h}_{II,t}$ in (9.53) and (9.54) by $\mathcal{Y}^{t}_{II,t}$. Together, these yield (9.47), and hence (ii), now with joint convergence with (i) and (iii) and the asserted independence. □

10. Remaining proofs

Proof of Lemma 3.9. By Lemma 9.8,

$$e^{-\lambda_1 t} b\cdot\mathcal{X}(t) \overset{\mathrm{a.s.}}{\to} (b\cdot v_1)W \qquad \text{as } t\to\infty, \tag{10.1}$$

which shows that $b\cdot\mathcal{X}(t)\to\infty$ a.s. when $W>0$. The remaining properties are immediate, observing that $\sup_{t\le T} b\cdot\mathcal{X}(t)<\infty$ for every finite T. □

The next lemma extends a result by Athreya and Karlin [7], [9, Theorem V.7.3].

Lemma 10.1. *Assume $b\cdot v_1>0$. As $z\to\infty$, (with $\ln 0=-\infty$)*

$$\tau_b(z) - \frac{1}{\lambda_1}\ln z \overset{\mathrm{a.s.}}{\to} -\frac{1}{\lambda_1}\bigl(\ln W + \ln(b\cdot v_1)\bigr).$$

Proof. Let $\mathcal{E}$ be the event $b\cdot\mathcal{X}(t)\to\infty$. On $\mathcal{E}$, $\tau_b(z)<\infty$ for all z and $\tau_b(z)\to\infty$ as $z\to\infty$. Hence (10.1) implies that, a.s. on $\mathcal{E}$,

$$e^{-\lambda_1\tau_b(z)} b\cdot\mathcal{X}\bigl(\tau_b(z)\bigr)\to(b\cdot v_1)W \qquad \text{as } z\to\infty. \tag{10.2}$$

By the right-continuity of $\mathcal{X}(t)$, $b\cdot\mathcal{X}\bigl(\tau_b(z)\bigr)\ge z$. On the other hand, if $\varepsilon(z)=1/z$, say, and z is so large that $\tau_b(z)>\varepsilon(z)$, then $b\cdot\mathcal{X}(\tau_b(z)-\varepsilon(z))<z$, and, again by (10.1),

$$e^{-\lambda_1\tau_b(z)} b\cdot\mathcal{X}(\tau_b(z)-\varepsilon(z))\to(b\cdot v_1)W \qquad \text{as } z\to\infty. \tag{10.3}$$

Combining (10.2) and (10.3) we find that, a.s. on $\mathcal{E}$,

$$e^{-\lambda_1\tau_b(z)} z\to(b\cdot v_1)W \qquad \text{as } z\to\infty. \tag{10.4}$$

On the complement of $\mathcal{E}$, $W=0$ a.s. by Lemma 3.9 and $\tau_b(z)=\infty$ for large z; hence (10.4) holds trivially. In other words, (10.4) holds a.s. Taking logarithms, we obtain the lemma. □

Lemma 10.2. *As $t\to\infty$, $e^{-\lambda_1 t/2}\sup_{s\le t}|\Delta\mathcal{X}(s)| \overset{\mathrm{a.s.}}{\to} 0$.*

Proof. Let $\Delta^*\mathcal{X}(t) := \sup_{s\le t}|\Delta\mathcal{X}(s)|$ and let $M>0$. Define

$$Z(t) := \sum_{i=1}^{q}\sum_{k:\tau_{ik}\le t}\mathbf{1}[|\Delta\mathcal{X}(\tau_{ik})|>M], \tag{10.5}$$

the number of jumps larger than M until time t. Clearly,

$$\mathbb{P}\bigl(\Delta^*\mathcal{X}(t)>M\bigr)=\mathbb{P}\bigl(Z(t)\ge 1\bigr)\le\mathbb{E}\,Z(t).$$

Moreover, applying Lemma 9.3 with $f_1(s) = f_2(s) = 1$ and $g(x) = \mathbf{1}[|x| > M]$ to the inner sum in (10.5) and using (9.11),

$$\mathbb{E}\, Z(t) = \sum_{i=1}^{q} \int_0^t \mathbb{P}(|\xi_i| > M) a_i \,\mathbb{E}\, \mathcal{X}_i(s)\, ds = \sum_{i=1}^{q} \mathbb{P}(|\xi_i| > M) \cdot O(e^{\lambda_1 t}).$$

Now let $\varepsilon > 0$ and let $\mathcal{E}_n$ be the event $\sup_{t\in[n,n+1]} e^{-\lambda_1 t/2} \Delta^* \mathcal{X}(t) > \varepsilon$. Then, taking $M = \varepsilon e^{\lambda_1 n/2}$,

$$\mathbb{P}(\mathcal{E}_n) \le \mathbb{P}\bigl(\Delta^*\mathcal{X}(n+1) > M\bigr) \le \mathbb{E}\, Z(n+1) \le C \sum_{i=1}^{q} \mathbb{P}(|\xi_i| > M) e^{\lambda_1(n+1)}$$

$$= C \sum_{i=1}^{q} \mathbb{P}(\varepsilon^{-2}|\xi_i|^2 > e^{\lambda_1 n}) e^{\lambda_1(n+1)}.$$

Summing over n we find

$$\sum_n \mathbb{P}(\mathcal{E}_n) \le C_1 \sum_{i=1}^{q} \mathbb{E} \sum_n e^{\lambda_1(n+1)} \mathbf{1}[e^{\lambda_1 n} < \varepsilon^{-2}|\xi_i|^2] \le C_2 \sum_{i=1}^{q} \mathbb{E}\, \varepsilon^{-2}|\xi_i|^2 < \infty,$$

and the Borelli–Cantelli lemma completes the proof. □

Lemma 10.3. *Conditioned on $W > 0$, $z^{-1/2} \sup_{s\le\tau_b(z)} |\Delta\mathcal{X}(s)| \overset{\text{a.s.}}{\to} 0$ as $z \to \infty$.*

Proof. Conditioned on $W > 0$, we have $\tau_b(z) \overset{\text{a.s.}}{\to} \infty$ by Lemma 3.9 and thus $e^{-\lambda_1\tau_b(z)/2} \sup_{s\le\tau_b(z)} |\Delta\mathcal{X}(s)| \overset{\text{a.s.}}{\to} 0$ by Lemma 10.2. Moreover, by Lemma 10.1 (or (10.4)), $z^{-1/2} e^{\lambda_1\tau_b(z)/2} \overset{\text{a.s.}}{\to} (b\cdot v_1)^{-1/2} W^{-1/2}$, and the result follows by multiplication. □

Proof of Theorem 3.10. We condition on $W > 0$, which by Lemma 9.7 is the same as essential non-extinction, and let $z \to \infty$.

First, $\tau_b(z) \to \infty$ and thus by Lemma 9.8 $e^{-\lambda_1\tau_b(z)} \mathcal{X}\bigl(\tau_b(z)\bigr) \overset{\text{a.s.}}{\to} W v_1$. Dividing by (10.4), we find (3.4).

Secondly, let $t = \lambda_1^{-1} \ln z$ and $x = \tau_b(z) - t$. Then $t \to \infty$ and, by Lemma 10.1, $x \to x_0 := -\lambda_1^{-1}\bigl(\ln W + \ln(b\cdot v_1)\bigr)$ a.s. Thus Theorem 3.1(i) yields

$$e^{-\lambda_1 t/2} P_I \mathcal{X}(t+x) \overset{\text{d}}{\to} e^{\lambda_1 x_0/2} W^{1/2} U_I(x_0) = (b\cdot v_1)^{-1/2} U_I(x_0).$$

Here x_0 is random, but U_I is a stationary process independent of W and thus of x_0, and thus $U_I(x_0) \overset{\text{d}}{=} U_I(0)$. Hence (i) follows with $V_I = U_I(0)$.

Next, let $t = \lambda_1^{-1} \ln z$ and $x = \tau_b(z)/t$. Thus Lemma 10.1 yields $xt - t \to -\lambda_1^{-1}\bigl(\ln W + \ln(b\cdot v_1)\bigr)$ and $x \to 1$ a.s. Hence Theorem 3.1(ii) implies for $\lambda \in \Lambda_{II}$, and thus $\operatorname{Re}\lambda = \lambda_1/2$,

$$t^{-(k+1/2)} e^{-\lambda_1 t/2 - \mathrm{i}\operatorname{Im}\lambda x t} Q_{\lambda,k} P_\lambda \mathcal{X}\bigl(\tau_b(z)\bigr) \overset{\text{d}}{\to} (b\cdot v_1)^{-1/2} U_{\lambda,k}(1). \qquad (10.6)$$

If $\lambda = \lambda_1/2$, (10.6) yields the limit in (ii) with $V_{\lambda,k} = \lambda_1^{-k-1/2} U_{\lambda,k}(1)$. If $\operatorname{Im}\lambda \neq 0$, we have a further factor $e^{-\mathrm{i}\operatorname{Im}\lambda xt}$ of modulus 1 on the left-hand side of (10.6). In this case, however, the distribution of $U_{\lambda,k}(1)$ is symmetric complex Gaussian (Remark 3.6), and it follows that (ii) follows in this case too. (Consider subsequences where $e^{-\mathrm{i}\operatorname{Im}\lambda t}$ converges.)

Similarly, (iii) follows from Theorem 3.1(iii), with $\check{W}_{\lambda,k} := \lambda_1^{-k} W^{-\lambda/\lambda_1} W_{\lambda,k}$.

Next, $|b\cdot\mathcal{X}(\tau_b(z)) - z| \leq |b||\Delta\mathcal{X}(\tau_b(z))|$ and thus $z^{-1/2}\big(b\cdot\mathcal{X}(\tau_b(z)) - z\big) \overset{\text{a.s.}}{\to} 0$ by Lemma 10.3. We multiply this by $(b\cdot v_1)^{-1}v_1$, which yields (iv) if we write $\mathcal{X} = \sum_\lambda P_\lambda\mathcal{X}$ and observe that for every $v \in \operatorname{Im} P_{\lambda_1} = E_{\lambda_1}$, $(b\cdot v/b\cdot v_1)v_1 = v$. (v) is a consequence of (iv) and $\mathcal{X} = \sum_\lambda P_\lambda\mathcal{X}$.

The result for urn processes follows by the embedding argument by Athreya and Karlin [8] discussed in the introduction. □

Proof of Corollary 3.11. By assumption, $\Lambda = \Lambda_I \cup \{\lambda_1\}$. Hence Theorem 3.10(v) yields, with some $R(z) \overset{\text{a.s.}}{\to} 0$,

$$z^{-1/2}\Big(\mathcal{X}\big(\tau_b(z)\big) - \frac{z}{b\cdot v_1}v_1\Big) = z^{-1/2}\Big(I - \frac{v_1 b'}{b\cdot v_1}\Big)P_I\mathcal{X}\big(\tau_b(z)\big) + R(z).$$

Theorem 3.10(i) shows that this converges to the Gaussian limit

$$(b\cdot v_1)^{-1/2}\Big(I - \frac{v_1 b'}{b\cdot v_1}\Big)V_I$$

with the covariance matrix (3.7), by (3.5) and (2.15); note that $\big(I - (b\cdot v_1)^{-1}v_1 b'\big)P_I = I - (b\cdot v_1)^{-1}v_1 b'$ because $P_I = I - P_{\lambda_1} = I - v_1 u_1'$ and $\big(I - (b\cdot v_1)^{-1}v_1 b'\big)v_1 = 0$. □

Proof of Corollary 3.12. This is similar to the proof of the preceding corollary; we now use Theorem 3.10(v), (i) and (ii), and find the Gaussian limit

$$(b\cdot v_1)^{-1/2}\Big(I - \frac{v_1 b'}{b\cdot v_1}\Big)\sum_{\lambda\in\Lambda_{II}} V_{\lambda,d}$$

(where $V_{\lambda,d} = 0$ when $d_\lambda < d$), which yields (3.8) by (3.6) and (2.17). □

Proof of Corollary 3.13. By Theorem 3.10(v),

$$Y_b(z) - \Big(I - \frac{v_1 b'}{b\cdot v_1}\Big)\sum_{\lambda\neq\lambda_1}(\ln z)^{-d} z^{-\operatorname{Re}\lambda_2/\lambda_1} P_\lambda\mathcal{X}\big(\tau_b(z)\big) \overset{\text{a.s.}}{\to} 0.$$

By Theorem 3.10(i), (ii) and (iii) (with $k = d_\lambda$), the terms in the sum with $\lambda \notin \tilde{\Lambda}'_{III} := \{\lambda : \operatorname{Re}\lambda = \operatorname{Re}\lambda_2,\ d_\lambda = d\}$ tend to 0 in probability; this can be improved to almost surely for $\lambda \in \Lambda_{III}$ by Theorem 3.10(iii) and for $\lambda \in \Lambda_I \cup \Lambda_{II}$ by the argument in the proof of Lemma 9.8; we omit the details. Moreover, for $\lambda \in \tilde{\Lambda}'_{III}$, Theorem 3.10(iii) implies

$$(\ln z)^{-d} z^{-\operatorname{Re}\lambda_2/\lambda_1} P_\lambda\mathcal{X}\big(\tau_b(z)\big) - z^{\mathrm{i}\operatorname{Im}\lambda/\lambda_1}(b\cdot v_1)^{-\lambda/\lambda_1}\check{W}_{\lambda,k} \overset{\text{a.s.}}{\to} 0.$$

Hence,

$$Y_b(z) - \sum_{\lambda\in\tilde\Lambda'_{III}} z^{\mathrm{i}\,\mathrm{Im}\,\lambda/\lambda_1}\widetilde{W}_{b,\lambda} \overset{\text{a.s.}}{\to} 0,$$

where $\widetilde{W}_{b,\lambda} := \bigl(I-\frac{v_1b'}{b\cdot v_1}\bigr)(b\cdot v_1)^{-\lambda/\lambda_1}\check{W}_{\lambda,k}$, and (3.9) follows with $W_{b,\lambda}=\widetilde{W}_{b,\lambda}$ when $\mathrm{Im}\,\lambda=0$ and $W_{b,\lambda}=2\widetilde{W}_{b,\lambda}$ when $\mathrm{Im}\,\lambda>0$. Tightness follows, as does (3.10) in the special case when $\Lambda_{III}=\{\lambda_2\}$ with λ_2 real. □

Proof of Theorem 3.16. As said in Section 3, we change the rules and add an additional dummy ball of a new type $q+1$ whenever a ball splits; dummy balls have $a_{q+1}=0$ and $\xi_{q+1}=0$ and thus never split. This does not affect the process of the balls of types $1,\dots,q$, but adds a count of the number of splits. Note that (A1)–(A6) hold for the modified process too. (However, it is not irreducible even if the original process is; this is our main reason for not assuming irreducibility in this paper.)

We write $\widetilde{}$ over the symbols to denote the modified process and various quantities defined for it. Writing vectors and matrices in block form, corresponding to a split $\mathbb{R}^q\times\mathbb{R}$, we see that, for $i=1,\dots,q$,

$$\widetilde{\xi}_i = \begin{pmatrix}\xi_i\\1\end{pmatrix}, \qquad \widetilde{a}=\begin{pmatrix}a\\0\end{pmatrix}, \qquad \widetilde{A}=\begin{pmatrix}A & 0\\ a' & 0\end{pmatrix}.$$

Consequently, the eigenvalues $\widetilde{\Lambda}=\Lambda\cup\{0\}$, in particular, $\tilde\lambda_1=\lambda_1$. It is easily verified that $\widetilde{A}$ has the corresponding eigenvectors

$$\widetilde{u}_1=\begin{pmatrix}u_1\\0\end{pmatrix}, \qquad \widetilde{v}_1=\begin{pmatrix}v_1\\ \lambda_1^{-1}\end{pmatrix}. \tag{10.7}$$

Finally, we choose $b=\binom{0}{1}$, which means that $\tau_b(n)$ is the first time we have n dummy balls, i.e. the n:th split time τ_n, and thus $\widetilde{X}_n=\widetilde{\mathcal{X}}\bigl(\tau_b(n)\bigr)$. By (10.7), we have $b\cdot\widetilde{v}_1=\lambda_1^{-1}$.

We may now apply Theorem 3.10 to $\widetilde{X}_n=\widetilde{\mathcal{X}}\bigl(\tau_b(n)\bigr)$, and the result follows from (3.4). □

Lemma 10.4. (i) *We have* $\psi(s,A)v_1=v_1$.

(ii) *For* $s\ge 1$ *and* $\lambda\in\Lambda$,

$$\psi(s,A)P_\lambda=\begin{cases}P_{\lambda_1}, & \lambda=\lambda_1,\\ \frac{s^{d_\lambda}}{d_\lambda!}e^{\lambda s}\bigl(I-\lambda^{-1}\lambda_1 v_1 a'\bigr)N_\lambda^{d_\lambda}P_\lambda+O(s^{d_\lambda-1}e^{\lambda s}), & 0<\mathrm{Re}\,\lambda<\lambda_1,\\ O(s^{d_\lambda+1}), & \mathrm{Re}\,\lambda\le 0.\end{cases}$$

Proof. By (3.11) and (3.12),

$$\phi(s,A)v_1=\int_0^s e^{tA}v_1\,dt=\int_0^s e^{t\lambda_1}v_1\,dt=\lambda_1^{-1}\bigl(e^{s\lambda_1}-1\bigr)v_1,$$

and thus

$$\psi(s,A)v_1=e^{s\lambda_1}v_1-\lambda_1v_1a'\lambda_1^{-1}\bigl(e^{s\lambda_1}-1\bigr)v_1=e^{s\lambda_1}v_1-v_1\bigl(e^{s\lambda_1}-1\bigr)=v_1.$$

Hence, (2.7) implies $\psi(s,A)P_{\lambda_1}=P_{\lambda_1}$.

Similarly, by (2.8), when $\operatorname{Re}\lambda > 0$ and $s \ge 1$,

$$\begin{aligned}\phi(s,A)P_\lambda &= \int_0^s e^{tA}P_\lambda\,dt = \sum_{j=0}^{d_\lambda}\frac{N_\lambda^j}{j!}P_\lambda\int_0^s t^j e^{\lambda t}\,dt \\ &= (d_\lambda!)^{-1}\lambda^{-1}s^{d_\lambda}e^{\lambda s}N_\lambda^{d_\lambda}P_\lambda + O(s^{d_\lambda-1}e^{\lambda s}).\end{aligned} \tag{10.8}$$

For $\operatorname{Re}\lambda \le 0$, the same argument yields $\phi(s,A)P_\lambda = O(s^{d_\lambda+1})$. The result follows from this, (3.12) and (2.8). □

Proof of Theorem 3.17. We continue the argument from the proof of Theorem 3.16. We have, see (2.13) and (2.14), for $i = 1,\dots,q$,

$$\widetilde{B}_i = \mathbb{E}\begin{pmatrix}\xi_i\\1\end{pmatrix}\begin{pmatrix}\xi_i' & 1\end{pmatrix} = \mathbb{E}\begin{pmatrix}\xi_i\xi_i' & \xi_i\\ \xi_i' & 1\end{pmatrix} = \begin{pmatrix}B_i & \mathbb{E}\,\xi_i\\ \mathbb{E}\,\xi_i' & 1\end{pmatrix},$$

and thus, using (2.14), (2.18) and $a\cdot v = 1$,

$$\widetilde{B} = \sum_{i=1}^{q+1}\widetilde{v}_{1i}\widetilde{a}_i\widetilde{B}_i = \sum_{i=1}^{q} v_{1i}a_i\widetilde{B}_i = \begin{pmatrix}B & \lambda_1 v_1\\ \lambda_1 v_1' & 1\end{pmatrix}. \tag{10.9}$$

Further, $\operatorname{Re}\tilde{\lambda}_2 = \max(\operatorname{Re}\lambda_2, 0) < \frac{1}{2}\tilde{\lambda}_1$, so we may apply Corollary 3.11 to $\widetilde{X}_n = \widetilde{\mathcal{X}}\big(\tau_b(n)\big)$, obtaining a Gaussian limit $\widetilde{V}_b$. Ignoring the dummy balls, we obtain $n^{-1/2}(X_n - \lambda_1 n v_1) \xrightarrow{\mathrm{d}} V$, with $V = (I,\ 0)\widetilde{V}_b$.

The covariance matrix of $\widetilde{V}_b$ is given by (3.7), with $\widetilde{\ }$ added everywhere. We have, by induction,

$$\widetilde{A}^n = \begin{pmatrix}A^n & 0\\ a'A^{n-1} & 0\end{pmatrix}, \qquad n \ge 1,$$

and thus

$$e^{s\widetilde{A}} = \sum_{n=0}^{\infty}\frac{s^n}{n!}\widetilde{A}^n = \begin{pmatrix}e^{sA} & 0\\ a'\phi(s,A) & 1\end{pmatrix}. \tag{10.10}$$

Hence

$$\begin{aligned}\Big(I - \frac{\widetilde{v}_1 b'}{b\cdot\widetilde{v}_1}\Big)e^{s\widetilde{A}} &= \left(\begin{pmatrix}I & 0\\ 0 & 1\end{pmatrix} - \lambda_1\begin{pmatrix}v_1\\ \lambda_1^{-1}\end{pmatrix}\begin{pmatrix}0 & 1\end{pmatrix}\right)e^{s\widetilde{A}} \\ &= \begin{pmatrix}I & -\lambda_1 v_1\\ 0 & 0\end{pmatrix}\begin{pmatrix}e^{sA} & 0\\ a'\phi(s,A) & 1\end{pmatrix} \\ &= \begin{pmatrix}\psi(s,A) & -\lambda_1 v_1\\ 0 & 0\end{pmatrix}.\end{aligned} \tag{10.11}$$

and thus, from (3.7), (10.9) and Lemma 10.4(i),

$$
\begin{aligned}
\mathbb{E}\, VV' &= \lambda_1 \int_0^\infty \bigl(\psi(s,A), \quad -\lambda_1 v_1\bigr)\, \widetilde{B}\, \bigl(\psi(s,A), \quad -\lambda_1 v_1\bigr)'\, e^{-\lambda_1 s} ds \\
&= \lambda_1 \int_0^\infty \Bigl(\psi(s,A) B \psi(s,A)' - \lambda_1^2 v_1 v_1' \psi(s,A)' - \lambda_1^2 \psi(s,A) v_1 v_1' \\
&\qquad\qquad + \lambda_1^2 v_1 v_1'\Bigr) e^{-\lambda_1 s} ds \\
&= \lambda_1 \int_0^\infty \Bigl(\psi(s,A) B \psi(s,A)' - \lambda_1^2 v_1 v_1'\Bigr) e^{-\lambda_1 s} ds,
\end{aligned}
$$

which yields (3.13). □

We need another lemma, which expresses the covariance matrix in (3.8) as a limit of the integral in (3.7) over finite intervals, suitably renormalized.

Lemma 10.5. *If* $\operatorname{Re}\lambda_2 = \lambda_1/2$ *and* $d := \max\{d_\lambda : \operatorname{Re}\lambda = \lambda_1/2\}$*, then, as* $t \to \infty$*,*

$$
t^{-2d-1} \int_0^t (I - P_{\lambda_1}) e^{sA} B e^{sA'} (I - P'_{\lambda_1}) e^{-\lambda_1 s}\, ds \to \Sigma_{II,d}, \tag{10.12}
$$

and thus, for any b *with* $b \cdot v_1 \neq 0$*,*

$$
\begin{aligned}
t^{-2d-1} \int_0^t \Bigl(I - \frac{v_1 b'}{b \cdot v_1}\Bigr) e^{sA} B e^{sA'} \Bigl(I - \frac{b v_1'}{b \cdot v_1}\Bigr) e^{-\lambda_1 s}\, ds& \\
\to \Bigl(I - \frac{v_1 b'}{b \cdot v_1}\Bigr) \Sigma_{II,d} \Bigl(I - \frac{b v_1'}{b \cdot v_1}\Bigr).&
\end{aligned} \tag{10.13}
$$

Proof. The left-hand side of (10.12) equals

$$
\begin{aligned}
&\sum_{\lambda,\mu \neq \lambda_1} t^{-2d-1} \int_0^t P_\lambda e^{s\lambda + sN_\lambda} B e^{s\mu + sN'_\mu} P'_\mu e^{-\lambda_1 s} ds \\
&\qquad = \sum_{\lambda,\mu \neq \lambda_1} t^{-2d-1} \sum_{k=0}^d \sum_{l=0}^d \int_0^t \frac{s^{k+l}}{k!\, l!} e^{(\lambda+\mu-\lambda_1)s} ds\, P_\lambda N_\lambda^k B (N'_\mu)^l P'_\mu .
\end{aligned}
$$

Here $\operatorname{Re}\lambda, \operatorname{Re}\mu \le \lambda_1/2$. For such λ and μ, the integral is $O(t^{2d})$ unless $\operatorname{Re}\lambda = \lambda_1/2$, $\mu = \bar\lambda$ and $k = l = d$, cf. the argument after (9.45). In the remaining case, the integral is $(2d+1)^{-1} d!^{-2} t^{2d+1}$, and (10.12) follows.

Multiplying (10.12) to the left by $\bigl(I - (b \cdot v_1)^{-1} v_1 b'\bigr)$ and to the right by the transpose, we obtain (10.13) because $\bigl(I - (b \cdot v_1)^{-1} v_1 b'\bigr) P_{\lambda_1} = 0$. □

Proof of Theorem 3.18. We argue as in the proof of Theorem 3.17, now applying Corollary 3.12. We have $\operatorname{Re}\tilde\lambda_2 = \operatorname{Re}\lambda_2 = \frac{1}{2}\tilde\lambda_1$. Moreover, it follows from (2.9) that $|e^{sA}(I - P_{\lambda_1})|$ grows as $s^d e^{\operatorname{Re}\lambda_2 s}$ as $s \to \infty$; it follows easily from (10.10) and (3.11) that $|e^{s\widetilde{A}}(I - \tilde P_{\lambda_1})|$ grows at the same rate, and thus $\tilde d = d$.

Hence, it follows from Corollary 3.12 that $(n \ln^{2d+1} n)^{-1/2}(X_n - n\lambda_1 v_1) \xrightarrow{\mathrm{d}} V := (I,\ 0)\widetilde{V}_b$, where $\widetilde{V}_b$ is Gaussian and, by (3.8) and Lemma 10.5,

$$\begin{aligned}\mathbb{E}\,\widetilde{V}_b\widetilde{V}_b' &= \lambda_1^{-2d}\Bigl(I - \frac{\widetilde{v}_1 b'}{b\cdot\widetilde{v}_1}\Bigr)\tilde{\Sigma}_{II,d}\Bigl(I - \frac{b\widetilde{v}_1'}{b\cdot\widetilde{v}_1}\Bigr)\\ &= \lambda_1^{-2d}\lim_{t\to\infty} t^{-2d-1}\int_0^t \Bigl(I - \frac{\widetilde{v}_1 b'}{b\cdot\widetilde{v}_1}\Bigr)e^{s\widetilde{A}}\widetilde{B}e^{s\widetilde{A}'}\Bigl(I - \frac{b\widetilde{v}_1'}{b\cdot\widetilde{v}_1}\Bigr)e^{-\lambda_1 s}\,ds.\end{aligned} \tag{10.14}$$

This implies, using (10.11), (10.9) and Lemma 10.4(i),

$$\begin{aligned}\mathbb{E}\,VV' &= \lambda_1^{-2d}\lim_{t\to\infty} t^{-2d-1}\int_0^t \bigl(\psi(s,A),\quad -\lambda_1 v_1\bigr)\,\widetilde{B}\,\bigl(\psi(s,A),\quad -\lambda_1 v_1\bigr)'\,e^{-\lambda_1 s}ds\\ &= \lambda_1^{-2d}\lim_{t\to\infty} t^{-2d-1}\int_0^t \Bigl(\psi(s,A)B\psi(s,A)' - \lambda_1^2 v_1 v_1'\Bigr)e^{-\lambda_1 s}ds.\end{aligned}$$

We write $\psi(s,A) = \sum_\lambda \psi(s,A)P_\lambda$ and use Lemma 10.4(ii); (3.14) follows.

For the final claim, note that

$$TP_\lambda N_\lambda^d = \lambda^{-1}\lambda_1 v_1 a' P_\lambda N_\lambda^d = \lambda^{-1}\lambda_1 v_1 \hat{a}' A P_\lambda N_\lambda^d = T_1 P_\lambda N_\lambda^d \qquad \square$$

Proof of Theorem 3.19. Follows directly by applying Corollary 3.13 to $\widetilde{X}_n$. $\square$

Proof of Theorem 3.21. We add dummy balls with $a_i = 0$, $i = q+1, \dots, 2q$, as described in Section 3, and argue as in the proofs of Theorems 3.16–3.18. For the modified process we now have, with vectors and matrices in block form corresponding to a split $\mathbb{R}^q \times \mathbb{R}^q$,

$$\widetilde{X}_n = \begin{pmatrix} X_n \\ N_n \end{pmatrix}, \qquad \widetilde{\xi}_i = \begin{pmatrix} \xi_i \\ \delta_i \end{pmatrix}, \qquad \widetilde{a} = \begin{pmatrix} a \\ 0 \end{pmatrix}, \qquad \widetilde{A} = \begin{pmatrix} A & 0 \\ D_a & 0 \end{pmatrix},$$

where $(\delta_i)_j = \delta_{ij}$ and $(D_a)_{ij} = a_i\delta_{ij}$. Again, the non-zero eigenvalues are the same: $\widetilde{\Lambda} = \Lambda \cup \{0\}$. $\widetilde{A}$ has the eigenvectors

$$\widetilde{u}_1 = \begin{pmatrix} u_1 \\ 0 \end{pmatrix}, \qquad \widetilde{v}_1 = \begin{pmatrix} v_1 \\ \lambda_1^{-1} D_a v_1 \end{pmatrix}.$$

We take $b = \binom{0}{\mathbf{1}}$, where $\mathbf{1}$ denotes the vector $(1, \dots, 1)'$, and note that $b \cdot \widetilde{v}_1 = \lambda_1^{-1} a \cdot v_1 = \lambda_1^{-1}$. The a.s. convergence follows from (3.4) and the Gaussian limits in (i) and (ii) follow by Corollaries 3.11 and 3.12.

For the explicit forms of the covariance matrices in (i) we use (3.7) and compute (omitting the details)

$$e^{s\widetilde{A}} = \sum_{n=0}^{\infty} \frac{s^n}{n!} \widetilde{A}^n = \begin{pmatrix} e^{sA} & 0 \\ D_a \phi(s,A) & I \end{pmatrix},$$

$$\Bigl(I - \frac{\widetilde{v}_1 b'}{b \cdot \widetilde{v}_1}\Bigr) e^{s\widetilde{A}} = \begin{pmatrix} \psi(s,A) & -\lambda_1 v_1 \mathbf{1}' \\ D_a(I - v_1 a')\phi(s,A) & I - D_a v_1 \mathbf{1}' \end{pmatrix}, \tag{10.15}$$

$$\widetilde{B} = \begin{pmatrix} B & AD_v \\ D_v A' & D_a D_v \end{pmatrix}. \tag{10.16}$$

Assuming $a = \mathbf{1}$, and thus $D_a = I$, the result now follows from (3.7) by simple calculations, using $D_v \mathbf{1} = v_1$ and $\phi(s,A)A = e^{sA} - I$.

For (ii), we use (3.8) and Lemma 10.5 and obtain (10.14) as in the proof of Theorem 3.18. We extract the leading terms in the integral in (10.14) using (10.15) and (10.16) together with Lemma 10.4 and (10.8), cf. the proof of Lemma 10.5. In the case $a = \mathbf{1}$, this yields, after some calculations, with T as in (3.14),

$$\mathbb{E} \begin{pmatrix} V \\ \widehat{V} \end{pmatrix} \begin{pmatrix} V' & \widehat{V}' \end{pmatrix} = \lambda_1^{-2d} \begin{pmatrix} I - T \\ \widehat{T} \end{pmatrix} \Sigma_{II,d} \begin{pmatrix} I - T' & \widehat{T}' \end{pmatrix}.$$

This yields the (co)variances. Since $A\widehat{T} P_\lambda N_\lambda^d = (I - T) P_\lambda N_\lambda^d$, $\lambda \in \Lambda_{II}$, it also implies that $\mathbb{E}(V - A\widehat{V})(V - A\widehat{V})' = 0$, and thus $V = A\widehat{V}$ a.s. □

Lemma 10.6. *For all $t \geq 0$,*

$$\mathbb{E} \sup_{s \leq t} |P_I \mathcal{X}(s)|^2 \leq C e^{\lambda_1 t}$$

and, if $\lambda \in \Lambda_{II}$ and $k \geq 0$,

$$\mathbb{E} \sup_{s \leq t} |Q_{\lambda,k} P_\lambda \mathcal{X}(s)|^2 \leq C(t+1)^{2k+1} e^{\lambda_1 t}.$$

Proof. It follows easily from (9.18), (9.11), (2.10) and (2.11) that, for $\lambda \in \Lambda_I$,

$$\mathbb{E} |P_\lambda \mathcal{X}(t)|^2 = O(e^{\lambda_1 t})$$

and, for $\lambda \in \Lambda_{II}$,

$$\mathbb{E} |Q_{\lambda,k} P_\lambda \mathcal{X}(t)|^2 = O\bigl((t+1)^{2k+1} e^{\lambda_1 t}\bigr).$$

For each u, $e^{A(u-t)} P_\lambda \mathcal{X}(t) = e^{Au} P_\lambda \mathcal{Y}(t)$ is a martingale, and thus by Doob's inequality

$$\mathbb{E} \sup_{u-1 \leq s \leq u} |P_\lambda \mathcal{X}(s)|^2 \leq C_1 \mathbb{E} \sup_{u-1 \leq s \leq u} |e^{A(u-s)} P_\lambda \mathcal{X}(s)|^2 \leq C_2 \mathbb{E} |P_\lambda \mathcal{X}(u)|^2,$$

and similarly for $Q_{\lambda,k} P_\lambda \mathcal{X}$, and the results follows by summing over all integers u less than $t + 1$. □

Proof of Theorem 3.22. We condition on $W > 0$.

(i): Let $t = \lambda_1^{-1} \ln z$. By Lemma 10.1,

$$\tau_b(xz) - t \to \phi(x) := \lambda_1^{-1}\bigl(\ln x - \ln W - \ln(b \cdot v_1)\bigr) \qquad \text{a.s.},$$

uniformly for $x \in K$ for any compact interval $K \subset (0, \infty)$. This together with Theorem 3.1(i) yields, in $D(0, \infty)$,

$$e^{-\lambda_1 t/2} P_I \mathcal{X}\bigl(\tau_b(xz)\bigr) \xrightarrow{\mathrm{d}} W^{1/2} e^{\lambda_1 \phi(x)/2} U_I\bigl(\phi(x)\bigr) = x^{1/2} (b \cdot v_1)^{-1/2} U_I\bigl(\phi(x)\bigr).$$

Since U_I is translation invariant and independent of W, the processes $U_I(\phi(x))$ and $U_I\bigl(\phi(x) - \phi(1)\bigr) = U_I(\lambda_1^{-1} \ln x)$ have the same distribution, and thus

$$z^{-1/2} P_I \mathcal{X}\bigl(\tau_b(xz)\bigr) = e^{-\lambda_1 t/2} P_I \mathcal{X}\bigl(\tau_b(xz)\bigr) \xrightarrow{\mathrm{d}} (b \cdot v_1)^{-1/2} V_I(x) \text{ in } D(0, \infty),$$

with $V_I(x) := x^{1/2} U_I(\lambda_1^{-1} \ln x)$. V_I is a continuous Gaussian process because U_I is, and (3.1) implies (3.17).

It remains to improve the result to convergence in $D[0, \infty)$, with $V_I(0) = 0$. For this, it suffices to show that for every $\varepsilon > 0$,

$$\limsup_{z \to \infty} \mathbb{P}\Bigl(\sup_{0 \le x \le h} z^{-1/2} |P_I \mathcal{X}\bigl(\tau_b(xz)\bigr)| > \varepsilon \Bigr) \to 0$$

as $h \to 0$, see e.g. [27, Proposition 2.4], and this is an easy consequence of Lemmas 10.6 and 10.1. Continuity of V_I at 0 follows, or is proved directly by standard methods [31, Theorem 3.23].

(ii): Convergence in $D(0, \infty)$ follows as in (i), now using Theorem 3.1(ii) and Lemma 10.1, observing as in the proof of Theorem 3.10(ii) that the constant factor $(b \cdot v_1 W)^{\mathrm{i} \operatorname{Im} \lambda / \lambda_1}$ of modulus one can be ignored. The convergence extends to $D[0, \infty)$ as in part (i), using Lemma 10.6(ii). □

Proof of Corollary 3.23. A simple consequence of Theorems 3.22 and 3.10(v); (3.18) follows from (3.2) and (2.16)–(2.17). □

Proof of Theorem 3.24. This follows from Corollary 3.23 by the arguments in the proofs of Theorems 3.17 and 3.18. In particular, the (co)variances in (3.19) follow easily using (10.11) and a change of variables; for (3.20) we observe that Corollary 3.23 shows that the covariance matrix depends on x and y only through the numerical factor $c(d, d, x, y)$, and the result follows by comparison with (3.14) (the case $x = y = 1$), where now $T = 2 v_1 a' P_{\lambda_2}$, together with (3.3). □

References

[1] D. Aldous, Asymptotic fringe distributions for general families of random trees. *Ann. Appl. Probab.* **1** (1991), no. 2, 228–266.

[2] D. Aldous, B. Flannery & J.L. Palacios, Two applications of urn processes. *Probab. Engineering Inform. Sci.* **2** (1988), 293–307.

[3] S. Asmussen, Almost sure behavior of linear functionals of supercritical branching processes. Trans. Amer. Math. Soc. **231** (1977), no. 1, 233–248.

[4] K.B. Athreya, Some results on multitype continuous time Markov branching processes. *Ann. Math. Statist.* **39** (1968), 347–357.

[5] K.B. Athreya, Limit theorems for multitype continuous time Markov branching processes. I. The case of an eigenvector linear functional. *Z. Wahrsch. Verw. Gebiete* **12** (1969), 320–332.
[6] K.B. Athreya, Limit theorems for multitype continuous time Markov branching processes. II. The case of an arbitrary linear functional. *Z. Wahrsch. Verw. Gebiete* **13** (1969), 204–214.
[7] K.B. Athreya & S. Karlin, Limit theorems for the split times of branching processes. *J. Math. Mech.* **17** (1967), 257–277.
[8] K.B. Athreya & S. Karlin, Embedding of urn schemes into continuous time Markov branching processes and related limit theorems. *Ann. Math. Statist.* **39** (1968), 1801–1817.
[9] K.B. Athreya & P.E. Ney, *Branching Processes.* Springer-Verlag, Berlin, 1972.
[10] A. Bagchi & A.K. Pal, Asymptotic normality in the generalized Pólya–Eggenberger urn model, with an application to computer data structures. *SIAM J. Algebraic Discrete Methods* **6** (1985), no. 3, 394–405.
[11] Z.D. Bai & F. Hu, Asymptotic theorems for urn models with nonhomogeneous generating matrices. *Stochastic Process. Appl.* **80** (1999), no. 1, 87–101.
[12] Z.D. Bai, F. Hu & L.-X. Zhang, Gaussian approximation theorems for urn models and their applications. *Ann. Appl. Probab.* **12** (2002), no. 4, 1149–1173.
[13] S. Bernstein, Nouvelles applications des grandeurs aléatoires presqu'indépendantes. (Russian) *Izv. Akad. Nauk SSSR Ser. Mat.* **4** (1940), 137–150.
[14] S. Bernstein, Sur un problème du schéma des urnes à composition variable. *C. R. (Doklady) Acad. Sci. URSS (N.S.)* **28** (1940), 5–7.
[15] P. Billingsley, *Convergence of Probability Measures.* Wiley, New York, 1968.
[16] H.-H. Chern & H.-K. Hwang, Phase changes in random m-ary search trees and generalized quicksort. *Random Structures Algorithms* **19** (2001), no. 3-4, 316–358.
[17] F. Eggenberger & G. Pólya, Über die Statistik verketteter Vorgänge. *Zeitschrift Angew. Math. Mech.* **3** (1923), 279–289.
[18] D.A. Freedman, Bernard Friedman's urn. *Ann. Math. Statist.* **36** (1965), 956–970.
[19] B. Friedman, A simple urn model. *Comm. Pure Appl. Math.* **2** (1949), 59–70.
[20] R. Gouet, Martingale functional central limit theorems for a generalized Pólya urn. *Ann. Probab.* **21** (1993), no. 3, 1624–1639.
[21] A. Gut, *Stopped Random Walks.* Springer-Verlag, New York, 1988.
[22] A. Gut & S. Janson, The limiting behaviour of certain stopped sums and some applications. *Scand. J. Statistics* **10** (1983), 281–292.
[23] H.-K. Hwang, Second phase changes in random m-ary search trees and generalized quicksort: convergence rates. Preprint, 2001.
[24] J. Jacod & A.N. Shiryaev, *Limit Theorems for Stochastic Processes.* Springer-Verlag, Berlin, 1987.
[25] S. Janson, Limit theorems for certain branching random walks on compact groups and homogeneous spaces. *Ann. Probab.* **11** (1983), 909-930.
[26] S. Janson, A functional limit theorem for random graphs with applications to subgraph count statistics, *Random Structures Algorithms* **1** (1990), 15–37.
[27] S. Janson, *Orthogonal Decompositions and Functional Limit Theorems for Random Graph Statistics.* Mem. Amer. Math. Soc., vol. 111, no. 534, American Mathematical Society, Providence, R.I., 1994.
[28] S. Janson, *Gaussian Hilbert Spaces.* Cambridge Univ. Press, Cambridge, 1997.
[29] S. Janson, Asymptotic degree distribution in random recursive trees. To appear. Available from `http://www.math.uu.se/~svante/`
[30] N.L. Johnson & S. Kotz, *Urn models and their application.* Wiley, New York, 1977.
[31] O. Kallenberg, *Foundations of modern probability.* 2nd ed., Springer-Verlag, New York, 2002.
[32] S. Karlin, *A First Course in Stochastic Processes.* Academic Press, New York, 1969.

[33] H. Kesten & B.P. Stigum, Additional limit theorems for indecomposable multidimensional Galton–Watson processes. *Ann. Math. Statist.* **37** (1966), 1463–1481.
[34] H. Kesten & B.P. Stigum, Limit theorems for decomposable multi-dimensional Galton–Watson processes. *J. Math. Anal. Appl.* **17** (1967), 309–338.
[35] S. Kotz, H. Mahmoud & P. Robert, On generalized Pólya urn models. *Statist. Probab. Lett.* 49 (2000), no. 2, 163–173.
[36] W. Lew & H.M. Mahmoud, The joint distribution of elastic buckets in multiway search trees. *SIAM J. Comput.* **23** (1994), no. 5, 1050–1074.
[37] T. Lindvall, Weak convergence of probability measures and random functions in the function space $D(0,\infty)$. *J. Appl. Probab.* **10** (1973), 109–121.
[38] H.M. Mahmoud, *Evolution of Random Search Trees.* Wiley, New York, 1992.
[39] H.M. Mahmoud, On rotations in fringe-balanced binary trees. *Inform. Process. Lett.* **65** (1998), no. 1, 41–46.
[40] H.M. Mahmoud, Urn models evolving by drawing multisets of balls. Preprint, 2000.
[41] H.M. Mahmoud, The size of random bucket trees via urn models. *Acta Inform.* **38** (2002), no. 11-12, 813–838.
[42] H.M. Mahmoud & B. Pittel, Analysis of the space of search trees under the random insertion algorithm. *J. Algorithms* **10** (1989), no. 1, 52–75.
[43] H.M. Mahmoud & R.T. Smythe, Asymptotic joint normality of outdegrees of nodes in random recursive trees. *Random Structures Algorithms* **3** (1992), no. 3, 255–266.
[44] H.M. Mahmoud, R.T. Smythe & J. Szymański, On the structure of random plane-oriented recursive trees and their branches. *Random Structures Algorithms* **4** (1993), no. 2, 151–176.
[45] A. Meir & J.W. Moon, Recursive trees with no nodes of out-degree one. *Congr. Numer.* **66** (1988), 49–62.
[46] K. Nomizu, *Fundamentals of linear algebra.* 2nd ed. Chelsea Publishing Co., New York, 1979.
[47] G. Pólya, Sur quelques points de la théorie des probabilités. *Ann. Inst. Poincaré* **1** (1931), 117–161.
[48] P. Protter, *Stochastic integration and differential equations. A new approach.* Applications of Mathematics, 21, Springer-Verlag, Berlin, 1990.
[49] V. Savkevitch, Sur le schéma des urnes à composition variable. *C. R. (Doklady) Acad. Sci. URSS (N.S.)* **28** (1940), 8–12.
[50] E. Seneta, *Nonnegative matrices and Markov chains.* 2nd ed. Springer-Verlag, New York, 1981.
[51] R.T. Smythe, Central limit theorems for urn models. *Stochastic Process. Appl.* **65** (1996), no. 1, 115–137.
[52] R.T. Smythe & W.F. Rosenberger, Play-the-winner designs, generalized Pólya urns, and Markov branching processes. In *Adaptive designs*, eds. N Flournoy & W.F. Rosenberger, IMS Lecture Notes Monogr. Ser. 25, Inst. Math. Statist., Hayward, CA, 1995, 13–22.
[53] L.J. Wei & S. Durham, The randomized play-the-winner rule in medical trials. *J. Amer. Statist. Assoc.* **73** (1978), issue 364, 840–843.
[54] L.J. Wei, R.T. Smythe, D.Y. Lin & T.S. Park, Statistical inference with data-dependent treatment allocation rules. *J. Amer. Statist. Assoc.* **85** (1990), issue 409, 156–162.

Department of Mathematics, Uppsala University, PO Box 480, S-751 06 Uppsala, Sweden
E-mail address: svante.janson@math.uu.se
URL: http://www.math.uu.se/~svante/