Adv. Appl. Prob. **50,** 504–542 (2018)
*doi:*10.1017/apr.2018.23

COAGULATION AND UNIVERSAL SCALING LIMITS FOR CRITICAL GALTON–WATSON PROCESSES†

GAUTAM IYER,* ** *Carnegie Mellon University*
NICHOLAS LEGER,*** *University of Houston*
ROBERT L. PEGO,* **** *Carnegie Mellon University*

Abstract

The basis of this paper is the elementary observation that the n-step descendant distribution of any Galton–Watson process satisfies a discrete Smoluchowski coagulation equation with multiple coalescence. Using this we obtain simple necessary and sufficient criteria for the convergence of scaling limits of critical Galton–Watson processes in terms of scaled family-size distributions and a natural notion of convergence of Lévy triples. Our results provide a clear and natural interpretation, and an alternate proof, of the fact that the Lévy jump measure of certain continuous-state branching processes (CSBPs) satisfies a generalized Smoluchowski equation. (This result was previously proved by Bertoin and Le Gall (2006).) Our analysis shows that the nonlinear scaling dynamics of CSBPs become *linear and purely dilatational* when expressed in terms of the Lévy triple associated with the branching mechanism. We prove a continuity theorem for CSBPs in terms of the associated Lévy triples, and use our scaling analysis to prove the existence of *universal* critical Galton–Watson processes and CSBPs analogous to Doeblin's 'universal laws'. Namely, these universal processes generate *all* possible critical and subcritical CSBPs as subsequential scaling limits. Our convergence results rely on a natural topology for Lévy triples and a continuity theorem for Bernstein transforms (Laplace exponents) which we develop in a self-contained appendix.

Keywords: Smoluchowski equation; multiple coalescence; Galton–Watson process; CSBP; universal law; Bernstein transform

2010 Mathematics Subject Classification: Primary 60J80
Secondary 37E20; 35Q70

1. Introduction

In 1873, Galton and Watson [44] undertook an investigation into the phenomenon of 'the decay of the families of men who occupied conspicuous positions in past times'. The problem, posed by Galton, was summarized by the Rev. H. W. Watson as follows:

> Suppose that at any instant all the adult males of a large nation have different surnames, it is required to find how many of these surnames will have disappeared in a given number of

† In memory of Jack Carr.

Received 3 November 2016; revision received 9 January 2018.

* Postal address: Department of Mathematical Sciences, Carnegie Mellon University, Pittsburgh, PA 15213, USA.
** Email address: gautam@math.cmu.edu
*** Postal address: Department of Mathematics, University of Houston, 4800 Calhoun Rd., Houston, TX 77004, USA. Email address: nleger@math.uh.edu
**** Email address: rpego@cmu.edu

> generations upon any hypothesis, to be determined by statistical investigations, of the law of male population.

Their analysis led to the eventual creation of a class of time-discrete, size-discrete branching processes now called Galton–Watson (GW) processes or Bienaymé–Galton–Watson processes, since key aspects of the topic were discovered thirty years earlier by I. J. Bienaymé (see [3] and [21]). The study of these and other branching processes has led to numerous interesting lines of research, and we refer the reader to [2] where the basic theory and applications are beautifully laid out.

Branching and coalescence are intrinsically connected concepts, and are in some sense dual. Indeed, following a tree from root to leaf leads to branching, and following it from leaf to root leads to coalescence. Mathematically, a deep connection between branching and coalescence was exposed by Kingman [23], [24] in his efforts to address questions of ancestry in population genetics. Loosely speaking, ancestral lineages branch as time proceeds forward, while the Kingman coalescent traces the size of groups or clusters of individuals with a common living ancestor as time evolves backwards. Kingman was particularly concerned with models that emerged from the early work of Moran and Wright in which the total population was fixed. His work was subsequently expanded to the study of a general class of exchangeable coalescent models, and a natural completion—the class of Λ-coalescents—was identified and studied by Pitman [36] and Sagitov [37]. We refer the reader to [5], [9], and [17] for an overview of the now-extensive literature in this area. For aspects particularly related to CSBPs, we refer the reader to [6], [7], [10], [14], [26], and [39].

In this paper we focus on classical GW processes and their continuum limits—continuous-state branching processes (CSBPs)—which are models in which populations may fluctuate. Our work was originally motivated by the observations of Bertoin and Le Gall contained in a series of papers that relate certain coalescence models to CSBPs; see [11]–[13]. In [13], the authors proved a striking result which shows that the Lévy jump measure associated to certain CSBPs satisfies a generalized Smoluchowski coagulation equation with multiple coalescence. One motivation for this work was to provide a simple and natural explanation for this fact.

Our starting point is an elementary observation showing that the n-step descendant distribution of the GW process satisfies a discrete Smoluchowski coagulation equation with multiple coalescence. We use this basic connection to establish simple criteria for the convergence of scaling limits of critical GW processes and study the coagulation dynamics of the Lévy jump measure of certain CSBPs. (As we consider only classical time-homogeneous GW processes, we make no connection with the large literature on branching in varying and random environments, except for the recent work of Bansaye and Simatos [4] where the authors developed general convergence criteria that compare closely with ours.)

Our scaling and limit analysis of the coagulation dynamics associated with critical GW processes leads to a simple yet striking observation: the *nonlinear scaling dynamics of CSBPs become linear and purely dilational* when expressed in terms of the Lévy triple that represents the branching mechanism. This observation is parallel to the results of Menon and Pego [33] that concern the dynamic scaling analysis of solvable Smoluchowski equations, and extends an analogy between CSBPs and infinite divisibility which has been evident since the work of Grimvall [19] in classifying all continuum limits of GW processes as CSBPs. Here we make use of the dilational representation to establish the existence of certain *universal* critical GW processes and CSBPs which generate *all* critical and subcritical CSBPs as subsequential scaling limits. These universal processes are analogous to Doeblin's 'universal laws' in classical probability theory; see [16, Section XVII.9]. Further, they supply a precise interpretation to a remark made by Grey [18] to the effect that a large class of 'critical and subcritical processes . . . do not seem to lend themselves to suitable scaling' which yields a well-defined limit.

1.1. Summary of results

The principal new results we prove in this paper include the existence of universal GW processes and the existence of universal CSBPs.

Theorem 1.1. *There exists a critical GW process $X_\star$ such that any (sub)critical CSBP that remains finite almost surely can be obtained as a subsequential scaling limit of $X_\star$.*

Theorem 1.2. *There exists a critical CSBP $Z_\star$ such that any (sub)critical CSBP that remains finite almost surely can be obtained as a subsequential scaling limit of $Z_\star$.*

These results follow naturally and directly from our analysis of discrete coalescence models in terms of the topology of Lévy triples that we introduce. The proof of Theorem 1.2, in particular, is based upon a new continuity theorem (Theorem 9.1) which establishes that convergence of (sub)critical CSBPs is equivalent to convergence of the Lévy triples which generate them. Our paper also contains a variety of other results concerning these issues. In the rest of this section we summarize the main results and organization of the paper.

1.1.1. *Time and size discrete coalescence and branching.* Sections 2–4 are devoted to the study of discrete-time, discrete-size coalescence and branching. We begin by describing (Section 2) a time-discrete Markov process C modeling sizes of clusters undergoing coalescence. (A more general class of processes of the type we study was introduced independently by Grosjean and Huillet [20], as discussed in Section 2 below.) We show (Proposition 2.2) that the backward equation of the process C is exactly a time-discrete analog of the well-known Smoluchowski coagulation equation [42], [43], a mean-field rate equation model of clustering. Moreover, we show that the one-dimensional distribution of coordinates of C has the same distribution as a GW process (see [20] for more extensive results concerning genealogies). This provides an elementary and fundamental connection between branching and coalescence, and shows that the nth generation descendant distribution of a critical GW process is itself a solution of the discrete Smoluchowski equation with multiple coalescence. This will be used extensively in the scaling analysis carried out in Sections 5–7.

We conclude our treatment of the discrete scenario by introducing the *branching mechanism* and the *Bernstein transform* for GW processes. We show (Proposition 4.1) that the discrete coagulation dynamics leads to an elegant (discrete-time) evolution equation for the Bernstein transform in terms of the branching mechanism. This parallels the evolution of the Bernstein transform of the size distribution in the Smoluchowski dynamics (see [13] and [22], or (5.3)), and is a direct analog of the backward equation for critical CSBPs that become extinct almost surely.

1.1.2. *Convergence criteria for scaling limits of critical GW processes.* Sections 5–7 concern the study of scaling limits of critical GW processes. We begin with a brief introduction to CSBPs (Section 5) and discuss Bertoin and Le Gall's striking result from [13] showing that the Lévy jump measure of certain CSBPs satisfies a special form of the Smoluchowski equation with multiple coalescence.

Using improvements in the theory of Bernstein functions [38] (known in probability as Laplace exponents), we obtain simple and precise criteria for the existence of scaling limits of critical GW processes expressed directly in terms of Lévy convergence of rescaled reproduction laws (Propositions 6.1–6.3). These convergence results provide a clear description of how the Lévy jump measure of certain CSBPs arises as the rescaled limit of n-step descendant distributions of GW processes, and explains how the generalized Smoluchowski equation arises naturally in [13]. The proofs make use of a continuity theorem for Bernstein transforms which appears to be little known. As we expect this to be of wider utility, we develop this theory separately in Appendix A. The well-known continuity theorems guarantee that weak-$\star$ convergence of probability measures is equivalent to pointwise convergence of their characteristic functions (or Laplace transforms). In our context, the continuity theorem shows that convergence of Lévy triples is equivalent to pointwise convergence of the Bernstein transforms. The proof is similar to that of the classical continuity theorems. However, since it is not readily available in the literature we prove it in Appendix A (Theorem A.2).

Finally, the convergence of scaling limits of general GW processes is a classical subject and has been studied by many authors. In Section 7 we compare our criteria to the classical convergence criteria provided by Grimvall [19] and the simplified criteria recently derived by Bansaye and Simatos [4] in connection with their more general investigation of scaling limits for GW processes in varying environments. We establish (Proposition 7.2) equivalence between these criteria and Lévy triple convergence in the critical case.

1.1.3. *Universality in GW processes and CSBPs* . Sections 8 and 9 are devoted to studying universality and proving Theorems 1.1 and 1.2. For GW processes, we use our simplified convergence results to show that there exists a GW process with a 'universal' family-size distribution (Theorem 8.1).

The proof that the scaling dynamics of CSBPs become linear and purely dilatational when expressed in terms of the Lévy triple associated with the branching mechanism comes down to a simple calculation (Proposition 9.1). We then show that the map from Lévy triple to solution is bicontinuous in a particular sense, proving a continuity theorem for CSBPs (Theorem 9.1). From this and our study of dilational dynamics for Lévy triples in Section 8, we infer the existence of universal CSBPs whose subsequential scaling limits yield all possible critical and subcritical CSBPs (Theorem 1.2 above, restated precisely as Theorem 9.2).

2. Coagulation equations and processes with multiple mergers

The Smoluchowski coagulation equation [42], [43] is a mean-field rate equation model of coalescence that may be used to describe a system of clusters that merge as time evolves to form larger clusters. The type of objects that comprise the clusters varies widely in applications—examples include smog particles, animals, and dark matter. Smoluchowski's equation governs the time evolution of the *cluster-size distribution*, under certain assumptions which usually include the assumption that only binary mergers are taken into account. Our aim in this section is to describe a natural and elementary Markov process that directly relates to a time-discrete Smoluchowski equation generalized to account for simultaneous mergers of any number of clusters.

The coalescence processes that we describe bear a close relation to GW processes that will be delineated in the following section. As indicated in the introduction, the coalescence process we describe is a time-homogeneous version of a class of processes introduced independently by Grosjean and Huillet [20]. In [20] the authors developed more extensively a correspondence between coalescence and branching processes in terms of genealogies and ancestral trees. Our use of the coalescence process in this paper is quite different, simply to indicate a clear relation between branching and the Smoluchowski equation, as preparation for studying scaling limits.

2.1. A discrete-time process modeling coalescence

Let $\hat{\pi}$ be a probability measure on $\mathbb{N}_0 := \mathbb{N} \cup \{0\}$. We consider a countably infinite collection of clusters, each with nonnegative integer size, which undergo coalescence at discrete time steps according to the following rules.

(R1) At each time-step, the existing clusters are collected into disjoint groups (which may be empty), and the clusters in each group merge to form the new collection of clusters. Empty groups create clusters of zero size.

(R2) The probability that a merger involves exactly k clusters is $\hat{\pi}(k)$. We refer to $\hat{\pi}$ as the *merger distribution*.

(R3) The sizes of each of the k clusters participating in a simultaneous k-merger are independent and identically distributed (i.i.d.).

An essential feature of this process is that the distribution of cluster sizes after a simultaneous k-merger is the same as the distribution of the sum of k independent copies of the cluster sizes before the merger.

An elementary construction of a coagulation process $C = (C_n)_{n\in\mathbb{N}_0}$ having the above features proceeds as follows. For each time $n \in \mathbb{N}_0$, C_n is a random function from $\mathbb{N}$ to $\mathbb{N}_0$, whose values $C_n(j)$ represent the size of a cluster in the ensemble. At the initial time $n = 0$, the random variables $C_0(j),\ j \in \mathbb{N}$, may, in general, be taken as i.i.d.; however, for the sake of comparison with GW processes, here we will always assume that

$$C_0(j) = 1 \quad \text{for all } j \in \mathbb{N}.$$

Given C_n, we define C_{n+1} as follows. Choose a random sequence $(M_{k,n})_{k\in\mathbb{N}_0}$ independent of C_m for $0 \le m \le n$, such that $M_{0,n} = 0$ and the increments

$$M_{k+1,n} - M_{k,n} \in \mathbb{N}_0$$

are i.i.d. with law $\hat{\pi}$. Define

$$C_{n+1}(j) = \sum_{i=1+M_{j-1,n}}^{M_{j,n}} C_n(i) \tag{2.1}$$

with the convention that the sum is 0 if $M_{j,n} = M_{j-1,n}$ (this corresponds to the creation of a new cluster of size 0). The quantity $M_{j+1,n} - M_{j,n}$ represents the number of clusters that simultaneously merge and combine their sizes to form a new cluster.

Note that C is a Markov process. Moreover, for each $n \ge 0$, the random variables $C_n(i),\ i \in \mathbb{N}$, are i.i.d. Thus, (2.1) guarantees that for any $j \in \mathbb{N}$, the random variable $C_{n+1}(j)$ is the sum of N i.i.d. copies of $C_n(j)$, where N is itself a random variable with distribution $\hat{\pi}$ and independent of C_n. This shows that the process C meets conditions (R1)–(R3) above.

We remark, however, that while the sequence $(C_n(1), C_n(2), \dots)$ represents the sizes of all clusters in the system, the individual coordinate functions $n \mapsto C_n(j)$ do not track the time evolution of the size of a particular cluster and need not form a Markov process.

Remark 2.1. In the $\hat{\pi}(0) = 0$ case, the partial sums $(\sum_{j=1}^{k} C_n(j))_{k\in\mathbb{N}}$ correspond to the marks of a *renewal process* with integer increments for each $n \ge 0$. The coagulation process C corresponds to a discrete-time type of *thinning* of these processes—marks disappear according

https://doi.org/10.1017/apr.2018.23 Published online by Cambridge University Press

to the rules that govern the process C. It was pointed out by Aldous [1] that independent thinning of renewal processes yields a classical Smoluchowski coagulation equation in continuous time. We find the condition $\hat{\pi}(0) > 0$ necessary, however, in order to produce the correspondence with critical GW processes in Section 3 below.

2.2. Evolution of the cluster-size distribution

Let ν_n denote the distribution of cluster sizes at step n in the coagulation process above. That is, $\nu_n(j)$ denotes the chance that any single cluster has size j after n time steps. Since the variables $\{C_n(i)\}_{i\in\mathbb{N}}$ are all identically distributed, we have

$$\nu_n(j) = \mathbb{P}\{C_n(i) = j\} \quad \text{for all } i \in \mathbb{N}. \tag{2.2}$$

Our first observation is that the n-step size distribution ν_n determines ν_{n+1} in a manner that naturally captures (R1)–(R3).

Proposition 2.1. *For any $n \geq 0$ and $j \in \mathbb{N}$, we have*

$$\nu_{n+1}(j) = \sum_{k\geq 0} \hat{\pi}(k)\nu_n^{*k}(j). \tag{2.3}$$

Here ν_n^{*k} denotes the kth convolution power of ν_n and is given by

$$\nu_n^{*k}(j) = \sum_{i_1,\dots,i_k\geq 0} \delta^j(i_1 + \cdots + i_k)\nu_n(i_1)\cdots\nu_n(i_k),$$

where δ^j denotes a Kronecker delta function:

$$\delta^j(k) = \begin{cases} 1, & j = k, \\ 0, & j \neq k. \end{cases}$$

By convention, we define $\nu_n^{*0}(j) = \delta^0(j)$.

Proof of Proposition 2.1. The proof follows directly from (2.1), independence of the $C_n(i)$ ($i \in \mathbb{N}$), and the fact that $M_{k+1,n} - M_{k,n}$ has distribution $\hat{\pi}$. □

Remark 2.2. In light of the assumption $C_0 \equiv 1$, the size distribution after one time step is exactly the merger distribution (i.e. $\nu_1 = \hat{\pi}$). This follows immediately from (2.3).

2.3. The discrete Smoluchowski equation

The size distribution of the process C naturally relates to a discrete-time Smoluchowski equation with multiple coalescence, as we now explain. The key idea is to express the dynamics in terms of clusters of nonzero size. For convenience, we will focus on the case when the merger distribution is critical, according to the following terminology.

Definition 2.1. We say the merger distribution $\hat{\pi}$ is *critical* if

$$\sum_{k\geq 0} k\hat{\pi}(k) = 1. \tag{2.4}$$

This means the expected number of clusters involved in a simultaneous merger is 1.

https://doi.org/10.1017/apr.2018.23 Published online by Cambridge University Press

Proposition 2.2. *If the merger distribution $\hat{\pi}$ is critical then the size distribution ν_n satisfies*

$$\nu_{n+1}(j) - \nu_n(j) = \sum_{k\geq 2} \hat{R}_k(\rho_n)\hat{I}_k(\nu_n, j), \qquad j > 0, \tag{2.5}$$

where

$$\rho_n := \mathbb{P}\{C_n(1) > 0\} = \sum_{j>0} \nu_n(j), \tag{2.6}$$

$$\hat{R}_k(\rho_n) := \sum_{l\geq k} \hat{\pi}(l)\binom{l}{k}\rho_n^k(1-\rho_n)^{l-k}, \tag{2.7}$$

$$\hat{I}_k(\nu_n, j) := \sum_{i_1,\dots,i_k>0} \bigg[\delta^j\bigg(\sum_{l=1}^k i_l\bigg) - \sum_{l=1}^k \delta^j(i_l)\bigg]\frac{\nu_n(i_1)}{\rho_n}\cdots\frac{\nu_n(i_k)}{\rho_n}. \tag{2.8}$$

Remark 2.3. Equations (2.5)–(2.8) are a discrete version of the Smoluchowski equations (see Section 5) and can be interpreted as follows. Observe that at time n, the chance that any single cluster has positive size is exactly ρ_n. Consequently, the chance that any simultaneous merger involves exactly k clusters of nonzero size is given by $\hat{R}_k(\rho_n)$. Now in the event of a simultaneous merger of k nonzero-sized clusters, the chance that a cluster of size j is created is exactly the positive term in (2.8). On the other hand, a cluster of size j can itself be involved in this simultaneous merger, leading to the creation of a larger cluster and the destruction of the cluster of size j. The chance that this happens is the negative term in (2.8). Consequently, $\hat{I}_k(\nu_n, j)$ is the expected change in frequency of clusters of size j, given that a simultaneous merger of k nonzero sized clusters occurs. Assuming independence and summing yields (2.5).

Proof of Proposition 2.2. Equation (2.5) follows when we restrict the size dynamics to clusters of nonzero size. Indeed, $\nu_n(j)/\rho_n$ is exactly the chance that a cluster has size j, conditioned on having nonzero size. Since $\hat{R}_k$ is the chance that exactly k such clusters merge, we must have

$$\begin{aligned}\nu_{n+1}(j) &= \sum_{k\geq 1} \hat{R}_k(\rho_n)\bigg(\frac{\nu_n}{\rho_n}\bigg)^{*k}(j)\\ &= \sum_{k\geq 1} \hat{R}_k(\rho_n) \sum_{i_1,\dots,i_k>0} \delta^j(i_1+\cdots+i_k)\frac{\nu_n(i_1)}{\rho_n}\cdots\frac{\nu_n(i_k)}{\rho_n}.\end{aligned} \tag{2.9}$$

Next we compute the expected number of clusters of size $j > 0$ involved in a given merger. On the one hand, this should be exactly the expected number of clusters involved in a merger, multiplied by $\nu_n(j)$. On the other hand, this can be computed by counting the number of clusters of size j in every cluster merger involving at least one cluster with nonzero size. This yields

$$\bigg(\sum_{k\geq 0} k\hat{\pi}(k)\bigg)\nu_n(j) = \sum_{k\geq 1} \hat{R}_k(\rho_n) \sum_{i_1,\dots,i_k>0} (\delta^j(i_1)+\cdots+\delta^j(i_k))\frac{\nu_n(i_1)}{\rho_n}\cdots\frac{\nu_n(i_k)}{\rho_n}. \tag{2.10}$$

Using the criticality assumption (2.4), subtracting (2.10) from (2.9), and observing that the $k = 1$ term cancels, we obtain (2.5). □

https://doi.org/10.1017/apr.2018.23 Published online by Cambridge University Press

Remark 2.4. Without the criticality assumption (2.4), the above proof shows that (2.5) should be replaced with

$$\nu_{n+1}(j) - \nu_n(j) = (\Xi - 1)\nu_n(j) + \sum_{k\geq 2} \hat{R}_k(\rho_n)\hat{I}_k(\nu_n, j), \qquad j > 0, \tag{2.11}$$

if $\Xi := \sum_{k\geq 0} k\hat{\pi}(k)$, the expected number of clusters involved in a simultaneous merger, is finite.

3. GW processes and coagulation dynamics

In this section we establish a direct and elementary connection between critical GW processes and the coalescence processes described in Section 2.

3.1. A brief introduction to GW processes

The GW process was introduced to model the gradual extinction of Victorian aristocratic family names, despite an increase in the general population. The original model supposes that family names are passed down only through male heirs, and that each heir reproduces independently and identically to be replaced by his son or sons in the subsequent generation.

To fix notation, let $\hat{\pi}$ denote the family-size distribution, meaning $\hat{\pi}(j)$ is the chance that each heir has exactly j sons, and let X_n denote the number of male descendants of a single individual after n generations. Presuming the descendants reproduce independently, the process $X = \{X_n\}$ is a time-homogeneous Markov process taking values in $\mathbb{N}_0$ with $X_0 = 1$. Further, the n-step transition probabilities

$$P_n(i, j) := \mathbb{P}\{X_n = j \mid X_0 = i\}$$

satisfy the convolution property

$$P_n(i_1 + i_2, j) = \sum_{k=0}^{j} P_n(i_1, j - k) P_n(i_2, k).$$

The standard construction of a GW process is to choose an array $\{\xi_{i,j}\}$ of i.i.d. random variables with distribution $\hat{\pi}$, and define

$$X_{n+1} = \sum_{k=1}^{X_n} \xi_{k,n+1}. \tag{3.1}$$

We refer the reader to [2] for a more complete introduction.

Remark 3.1. Equation (2.1) appears dual to (3.1) in a curious way: the number of terms in the sum in (2.1) is distributed according to $\hat{\pi}$, while it is the individual terms in (3.1) that are so distributed. The individual terms in (2.1) are generated by the accumulation formula, while it is the number of terms in (3.1) that is so generated.

3.2. Discrete coagulation dynamics of the descendant distribution

Next we establish a direct, explicit, and elementary connection between the GW process X and the coagulation process C. Namely, we show that X_n, the number of descendants of a single individual after n generations, has the same distribution as the cluster sizes produced by the process C after n steps. More extensive connections between genealogies and ancestral trees of these processes (and related time-inhomogenous processes) were established in [20].

https://doi.org/10.1017/apr.2018.23 Published online by Cambridge University Press

Proposition 3.1. *Let X be a GW process, and let*

$$\hat{\nu}_n(j) := \mathbb{P}\{X_n = j\} = P_n(1, j)$$

denote the distribution of X_n. Suppose that C is a coagulation process as in Section 2.1 having merger distribution $\hat{\pi}$ equal to the family-size distribution of X. Then, for each $n \geq 0$, X_n has the same distribution as $C_n(i)$, $i \in \mathbb{N}$. That is,

$$\hat{\nu}_n = \nu_n \quad \textit{for all } n \geq 0,$$

where ν_n is defined in (2.2). Consequently, if X is critical then $\hat{\nu}_n$ satisfies the discrete Smoluchowski equations (2.5)–(2.8).

Recall that a GW process $\{X_n\}$ is said to be *critical* if the expected number of descendants of a single individual is 1. Explicitly, this means that $\hat{\pi}$ satisfies (2.4), and it is well known that such processes become extinct almost surely.

Proof of Proposition 3.1. According to branching equation (3.1) and the fact that if $X_n = i$ then the distribution of the sum is an i-fold convolution of $\hat{\pi}$, naturally we have

$$\hat{\nu}_{n+1}(j) = \sum_{i=0}^{\infty} \hat{\nu}_n(i)\hat{\pi}^{*i}(j) = \sum_{i=0}^{\infty} P_n(1, i)P_1(i, j).$$

However, we also have another characterization, due to the Markov property. Namely,

$$\hat{\nu}_{n+1}(j) = \sum_{k\geq 0} P_1(1, k)P_n(k, j) = \sum_{k\geq 0} \hat{\pi}(k)\hat{\nu}_n^{*k}(j).$$

This equation has exactly the same form as (2.3). Since $\hat{\nu}_0 = \nu_0$, we conclude the proof using induction. □

In light of Proposition 3.1, we identify $\hat{\nu}_n$ with ν_n for the remainder of this paper. While the proof of Proposition 3.1 is a simple algebraic calculation, a natural interpretation can be obtained through the work of Kingman; see [23] and [24]. To see this, fix N large, $n \leq N$, and divide the population at generation N into clans that have a common living ancestor at generation $N - n$. The clans here play the same role as the coalescing clusters in Section 2. Indeed, as n increases we look further back in the ancestry for a common living ancestor, leading to the merger of clans. Finally, one can directly check that for a given n the distribution of clan sizes is exactly $\hat{\nu}_n$, and, hence, (2.5)–(2.8) is expected.

4. Bernstein transform of the discrete evolution equations

The beautiful, classical theory of GW processes is normally developed in terms of the generating function for the family-size distribution $\hat{\pi}$, given by

$$G(z) = \sum_{j=0}^{\infty} \hat{\pi}(j)z^j.$$

The generating function of the nth generation descendant distribution (ν_n) is then given by the nth functional iterate of G. It is well known and straightforward to derive, from (2.3), that the function

$$G_n(z) = \sum_{j=0}^{\infty} \nu_n(j)z^j$$

https://doi.org/10.1017/apr.2018.23 Published online by Cambridge University Press

satisfies $G_0(z) = z$ and

$$G_{n+1}(z) = G(G_n(z)), \qquad n \geq 0. \tag{4.1}$$

In order to simplify the study of continuum limits and to compare with CSBPs, however, we find it convenient to recast the formulas of the theory using a representation more closely related to Laplace exponents.

Definition 4.1. Given a GW process X, we define its *Bernstein transform* by

$$\hat{\varphi}_n(q) = \mathbb{E}(1 - \mathrm{e}^{-qX_n}) = \sum_{j \geq 1} \nu_n(j)(1 - \mathrm{e}^{-qj}).$$

Of course, this is closely related to the Laplace transform, and may also be expressed in terms of generating functions as

$$\hat{\varphi}_n(q) = 1 - G_n(\mathrm{e}^{-q}). \tag{4.2}$$

The function $\hat{\varphi}_n$ is a *Bernstein function* [38]—a nonnegative function whose derivative is a Laplace transform. The class of Bernstein functions has a number of convenient properties—for example, it is closed under composition and takes pointwise limits—and such functions have proved valuable in previous studies on coagulation dynamics (see, for example, [30], [32], and [33]). Our first objective of this section is to obtain a convenient expression, (4.3), for the evolution of the Bernstein transform of a GW process.

Proposition 4.1. *If $\hat{\varphi}$ is the Bernstein transform of a GW process X then*

$$\hat{\varphi}_{n+1}(q) - \hat{\varphi}_n(q) = -\hat{\Psi}(\hat{\varphi}_n(q)) \quad \textit{for all } q \geq 0, \tag{4.3}$$

where

$$\hat{\Psi}(s) := \sum_{j=0}^{\infty}(1-s)^j \hat{\pi}(j) - 1 + s = G(1-s) - 1 + s \tag{4.4}$$

in terms of the family-size distribution $\hat{\pi}$ and its generating function G.

Proof. By (4.1) and (4.2), we have

$$\hat{\varphi}_{n+1}(q) - \hat{\varphi}_n(q) = 1 - G(G_n(\mathrm{e}^{-q})) - \hat{\varphi}_n(q) = -\hat{\Psi}(\hat{\varphi}_n(q)). \qquad \square$$

Definition 4.2. We define the function $\hat{\Psi}$ in (4.4) to be the *branching mechanism* of the GW process X.

While the generating function has proved extremely useful in many contexts, the notion of a branching mechanism as defined above is better suited for our purposes. As we will demonstrate, it is strongly analogous to the branching mechanism of a CSBP, it governs convergence of the process in the continuum limit (Proposition 6.1), and it also provides the discrete analog of the Poissonian structure of the rate constants arising in [13] (see (4.6) below and also revisited later in Section 5).

We remark that the branching mechanism $\hat{\Psi}$ is only guaranteed to be defined for $s \in [0, 2]$. It is convex for $s \in [0, 1]$. In terms of expected family size, it can also be written in the following form.

https://doi.org/10.1017/apr.2018.23 Published online by Cambridge University Press

Lemma 4.1. *If the expected family size* $\Xi = \sum_{k\geq 0} k\hat{\pi}(k)$ *is finite then the branching mechanism of the GW process X satisfies*

$$\hat{\Psi}(s) = \sum_{j=2}^{\infty}((1-s)^j - 1 + js)\hat{\pi}(j) + (1-\Xi)s. \tag{4.5}$$

We now compute the rate constants $\hat{R}_k$ (defined in (2.7)) in terms of the branching mechanism.

Lemma 4.2. *Let X be a GW process with branching mechanism* $\hat{\Psi}$*. Then, for all* $k \geq 2$*, we have*

$$\hat{R}_k(\rho_n) = \frac{(-\rho_n)^k}{k!}\hat{\Psi}^{(k)}(\rho_n), \tag{4.6}$$

where ρ_n *is defined by (2.6) and* $\hat{R}_k$ *is defined by (2.7).*

Proof. Termwise differentiation of (4.4) leads to

$$\frac{(-1)^k}{k!}\hat{\Psi}^{(k)}(s) = \sum_{j\geq k}\hat{\pi}(j)\binom{j}{k}(1-s)^{j-k}, \qquad k \geq 2, \tag{4.7}$$

for $s \in (0, 2)$. Since $\rho_n \in (0, 1)$ for all n, we may substitute $s = \rho_n$ in (4.7) and this yields (4.6) for $k \geq 2$. □

5. CSBPs and the time-continuous Smoluchowski equation

In subsequent sections we will consider scaling limits of GW processes, focusing on the critical case. The limiting processes obtained will be a class of CSBPs, and we use this section to summarize relevant properties of CSBPs. We also take this opportunity to indicate similarities between the CSBPs and the discrete notions introduced in Section 4 that foreshadow results in Section 6. Finally, we describe Bertoin and Le Gall's result [13] relating CSBPs to the time-continuous Smoluchowski equation and compare it to the discrete version introduced in Section 2.

A CSBP consists of a two-parameter random process $(x, t) \mapsto Z_t(x) \in [0, \infty)$ for $t \geq 0$ and $x > 0$. For fixed x, the process $t \mapsto Z_t(x)$ is a Markov process with initial value $Z_0(x) = x$. For fixed t, the process $x \mapsto Z_t(x)$ is an increasing process with independent and stationary increments. The right-continuous version of this is a Lévy process with increasing sample paths. In particular, the process enjoys the branching property that, for all $t \geq 0$, the distribution of $Z_t(x + y)$ is the same as the distribution of the sum of independent copies of $Z_t(x)$ and $Z_t(y)$.

The structure of the process Z has a precise characterization via the Lamperti transform (see [15] and [28] or [25, Chapter 12]). That is, $t \mapsto Z_t(x)$ can be expressed as a subordinated Markov process with parent process $x + \bar{X}_t$, where $\bar{X}_t$ is a Lévy process starting from 0 with no negative jumps (i.e. $\bar{X}_t$ is either spectrally positive or a subordinator). More specifically, $Z_t(x) = x + \bar{X}_{\Theta(x,t)}$, where the process $t \mapsto \Theta(x, t)$ has nondecreasing sample paths and formally solves $\partial_t\Theta = x + \bar{X}_\Theta$ with $\Theta(x, 0) = 0$. (A derivation of an analog of this for GW processes can be found in Remark 7.1, below.) In this context, the Laplace exponent of $\bar{X}_t$, denoted by Ψ, is called the *branching mechanism* for $Z_t(x)$ and has Lévy–Khintchine representation

$$\Psi(q) = \frac{1}{2}a_0q^2 - aq - b + \int_{(0,\infty)}(\mathrm{e}^{-qx} - 1 + qx\,\mathbf{1}_{\{x<1\}})\,\mathrm{d}\pi(x),$$

https://doi.org/10.1017/apr.2018.23 Published online by Cambridge University Press

where $a_0, b \geq 0$, $a \in \mathbb{R}$, and π is a positive measure on $(0, \infty)$ satisfying the finiteness condition

$$\int_{(0,\infty)} (1 \wedge x^2)\pi(\mathrm{d}x) < \infty.$$

It is well known (see [18] and [25]) that a CSBP remains finite almost surely (or is *conservative*) if and only if

$$\int_{(0,1)} \frac{1}{|\Psi(q)|}\,\mathrm{d}q = \infty.$$

Clearly, the branching mechanism of a conservative CSBP must have $b = 0$.

A CSBP is subcritical, critical, or supercritical if we have $\Psi'(0^+) > 0$, $\Psi'(0^+) = 0$, or $\Psi'(0^+) < 0$, respectively. Thus, for a finite (sub)critical CSBP (by which we mean a critical or subcritical CSBP that remains finite almost surely), we must have $b = 0$ and

$$0 \leq \Psi'(0^+) = -a - \int_{[1,\infty)} x\,\mathrm{d}\pi(x). \tag{5.1}$$

Consequently, the branching mechanism of a finite (sub)critical CSBP takes the form

$$\Psi(q) = \frac{1}{2}a_0 q^2 + a_\infty q + \int_{(0,\infty)} \frac{(\mathrm{e}^{-qx} - 1 + qx)}{x}\,\mathrm{d}\mu(x) \quad \text{for some } a_0, a_\infty \geq 0, \tag{5.2}$$

together with the finiteness condition

$$\int_{(0,\infty)} (x \wedge 1)\,\mathrm{d}\mu(x) < \infty.$$

Here the measures μ and π are related by $x\,\mathrm{d}\pi = \mathrm{d}\mu$. The reason we introduce the measure μ is because with this notation Ψ' is the Bernstein function associated with the Lévy triple (a_0, a_∞, μ). Note that $a_\infty = \Psi'(0^+)$ and can be expressed in terms of a using (5.1).

We remark that (5.2) closely parallels the form of the discrete branching mechanism $\hat{\Psi}$ in Definition 4.2. Indeed, for critical GW processes, $\hat{\Psi}$ takes the form (4.5), which is obtained from (5.2) by dropping the quadratic term $\frac{1}{2}a_0 q^2$, setting $a_\infty = 0$ by the criticality condition, and using the approximation $\mathrm{e}^{-s} \approx 1 - s$. We will see later (Proposition 6.1) that the quadratic term (along with a linear term) reappears in the continuum limit.

Returning to the CSBP Z, the nature of the Lamperti transform forces the relation

$$\mathbb{E}(\mathrm{e}^{-qZ_t(x)}) = \mathrm{e}^{-x\varphi_t(q)},$$

where the spatial Laplace exponent φ solves the backward equation

$$\partial_t \varphi_t(q) = -\Psi(\varphi_t(q)) \quad \text{with initial data } \varphi_0(q) = q. \tag{5.3}$$

This is exactly the continuum analog of the discrete equation (4.3), and we will provide a natural and elementary proof of it in Proposition 6.2.

As the Laplace exponent of a subordinator, φ has the Lévy–Khintchine representation

$$\varphi_t(q) = b_t q + \int_{(0,\infty)} (1 - \mathrm{e}^{-qx})\,\mathrm{d}\bar{\nu}_t(x), \qquad q \geq 0,$$

https://doi.org/10.1017/apr.2018.23 Published online by Cambridge University Press

where $b_t \geq 0$ and $\bar{\nu}_t$ is a positive measure satisfying the finiteness condition

$$\int_{(0,\infty)} (1 \wedge x) \,\mathrm{d}\bar{\nu}_t(x) < \infty.$$

The quantities b_t and $\bar{\nu}_t$ are, respectively, the *drift coefficient* and the *Lévy jump measure* of the CSBP Z.

A striking result of Bertoin and Le Gall [13] can be used to show that the Lévy jump measure of a *critical* CSBP which becomes extinct almost surely satisfies a generalized type of Smoluchowski coagulation equation. Explicitly, using Proposition 3 of [13], we have

$$\partial_t \langle f, \bar{\nu}_t \rangle = \sum_{k \geq 2} R_k I_k(\bar{\nu}_t, f) \quad \text{for all } f \in C([0,\infty]), \tag{5.4}$$

where

$$I_k(\bar{\nu}, f) := \int_{(0,\infty)^k} \biggl(f(x_1 + \cdots + x_k) - \sum_{i=1}^{k} f(x_i) \biggr) \prod_{i=1}^{k} \frac{\mathrm{d}\bar{\nu}(x_i)}{\langle 1, \bar{\nu} \rangle}, \tag{5.5}$$

$$R_k := \frac{(-\rho_t)^k \Psi^{(k)}(\rho_t)}{k!}, \qquad \rho_t := \langle 1, \bar{\nu}_t \rangle. \tag{5.6}$$

Here

$$\langle f, \bar{\nu}_t \rangle := \int_{(0,\infty)} f(x) \,\mathrm{d}\bar{\nu}_t(x)$$

is the duality pairing.

Equation (5.4) has a natural interpretation as a coagulation model introduced by Smoluchowski [42], [43] generalized to account for multiple coalescence. To understand this, we interpret $\{\bar{\nu}_t \mid t \geq 0\}$ as a family of positive measures on $\mathbb{R}^+ = (0, \infty)$ representing the size distribution of clusters. Namely, $\bar{\nu}_t(a, b)$ denotes the expected number of clusters at time t that have size in the interval (a, b).

Fix $k \geq 2$ and consider the change in the cluster size distribution due to the simultaneous merger of k clusters. We assume that the merging clusters are i.i.d. with distribution proportional to $\bar{\nu}_t$. For $y \in (0, \infty)$, the merging of smaller clusters into a cluster of size y will result in an increase in the density of clusters of size y. This will happen at a rate proportional to

$$\int_{(0,y)} \frac{\mathrm{d}\bar{\nu}_t(x_1)}{\langle 1, \bar{\nu}_t \rangle} \int_{(0,y-x_1)} \frac{\mathrm{d}\bar{\nu}_t(x_2)}{\langle 1, \bar{\nu}_t \rangle} \cdots \int_{(0,y-\sum_{i=1}^{k-2} x_i)} \frac{\mathrm{d}\bar{\nu}_t(x_{k-1})}{\langle 1, \bar{\nu}_t \rangle} \frac{\mathrm{d}\bar{\nu}_t(y - x_1 - \cdots - x_{k-1})}{\langle 1, \bar{\nu}_t \rangle}.$$

On the other hand, the clusters of size y also combine with larger clusters resulting in a decrease in the density of clusters of size y. This will happen at a rate proportional to

$$k \frac{\mathrm{d}\bar{\nu}_t(y)}{\langle 1, \bar{\nu}_t \rangle}.$$

Thus, for any test function $f \in C([0, \infty])$, the rate at which the simultaneous merger of k clusters affects the moment $\langle f, \bar{\nu}_t \rangle$ is proportional to the difference of the above two terms integrated against f. Changing variables, we see that this is exactly $I_k(f, \bar{\nu}_t)$ and, hence, the rate at which the simultaneous merger of k clusters affects $\langle f, \bar{\nu}_t \rangle$ is proportional to $I_k(f, \bar{\nu}_t)$. Summing over k and multiplying by proportionality constants explains how (5.4) models coalescence.

https://doi.org/10.1017/apr.2018.23 Published online by Cambridge University Press

In general, the rate constants R_k appearing in (5.4) can be chosen arbitrarily. In the context of CSBPs, the R_k have a special Poissonian structure given by (5.6). One of the main motivations of our exposition is to provide a clear account of the meaning of the measure $\bar{\nu}_t$ in this context, the precise way it arises in the continuum limit, and how it comes to be governed by coagulation dynamics with the indicated rates.

Precisely, we will show that for a finite (sub)critical CSBP, the Lévy jump measures $\bar{\nu}_t$ arise as the scaling limits of the nth generation descendant distributions of rescaled critical GW processes ν_n. To briefly explain the main idea, recall that due to Proposition 3.1, ν_n satisfies (2.5). This is, of course, simply a discrete version of (5.4). Indeed, given a sequence $\{f(j)\}_{j\in\mathbb{N}}$, we multiply (2.5) by $f(j)$ and sum over j to obtain

$$\sum_{j>0} f(j)(\nu_{n+1}(j) - \nu_n(j)) = \sum_{k\geq 2} \hat{R}_k(\rho_n)\hat{I}_k(\nu_n, f), \tag{5.7}$$

where

$$\hat{I}_k(\nu_n, f) := \sum_{i_1,\dots,i_k>0} \left[f\left(\sum_{l=1}^{k} i_l\right) - \sum_{l=1}^{k} f(i_l)\right] \prod_{l=1}^{k} \frac{\nu_n(i_l)}{\rho_n}.$$

Moreover, by Lemma 4.2, the rate constants can be obtained from the discrete branching mechanism $\hat{\Psi}$ (Definition 4.2) in exactly the same manner as (5.6).

Thus, after rescaling the GW processes correctly, it is only natural to expect convergence of the rescaled branching mechanisms of the GW process to the branching mechanism of the CSBP, and convergence of the rescaled descendant distributions ν_n to the Lévy jump measure ν_t. Moreover, the Lévy jump measure ν_t should satisfy (5.4) if the limit is critical, and (5.4) with an additional damping term (analogous to (2.11)) if the limit is subcritical. We prove this in Proposition 6.1 and Corollaries 6.1 and 6.2 below.

6. Scaling limits of critical GW processes

In this section we study scaling limits of critical GW processes using the discrete coagulation dynamics developed above.

We establish necessary and sufficient criteria for convergence of the discrete branching mechanisms (Proposition 6.1), convergence of the Bernstein transforms (Proposition 6.2), and of the rescaled GW processes themselves (Proposition 6.3), in terms of a type of weak convergence of the reproduction laws alone. Moreover, we show (Corollaries 6.1 and 6.2) that the Lévy jump measure of the limiting CSBP satisfies a generalized (damped) Smoluchowski equation. The precise notion of convergence is naturally associated with continuity theorems for Bernstein transforms, which we develop in Appendix A due to their independent interest.

6.1. Rescaled time-discrete dynamics

We begin by rescaling the coagulation model (2.5)–(2.8) (where $\hat{R}_k$ is defined by (2.7)). Let $h > 0$ be a grid size and $\tau > 0$ be a time step. We rescale the variables so that cluster sizes are integer multiples of h, and the merger of clusters happens on intervals of time τ. Further, in order to facilitate passing to the limit as $h, \tau \to 0$, we associate measures supported on the grid $h\mathbb{N}$ to the rescaled size distributions. Explicitly, we define

$$\nu_n^h = \frac{1}{h}\sum_{j\geq 1} \nu_n(j)\delta_{jh}, \qquad \pi_{h,\tau} = \frac{1}{\tau h}\sum_{j\geq 1} \hat{\pi}(j)\delta_{jh}. \tag{6.1}$$

Here δ_x denotes the Dirac measure centered at x.

https://doi.org/10.1017/apr.2018.23 Published online by Cambridge University Press

In the context of GW processes, the above corresponds to scaling the population by a factor of h and reproducing at times which are integer multiples of τ. That is, the rescaled process Y is given by

$$Y_{n\tau}(jh) = hX_n(j).$$

We will, however, postpone the discussion of rescaled GW processes to Section 6.4, and instead study the rescaled size distributions first.

Associated with (6.1), we denote the Bernstein transform of ν_n^h by

$$\varphi_n^h(q) = \int_{\mathbb{R}^+} (1 - \mathrm{e}^{-qx})\, \mathrm{d}\nu_n^h(x) = \frac{1}{h}\hat{\varphi}_n(hq).$$

We assume $\hat{\pi}$ is critical, and define a rescaled branching mechanism by

$$\Psi_{h,\tau}(q) := \int_{\mathbb{R}^+} ((1 - hq)^{x/h} - 1 + qx)\, \mathrm{d}\pi_{h,\tau}(x) = \frac{\hat{\Psi}(hq)}{\tau h}. \tag{6.2}$$

Note that $\Psi_{h,\tau}$ is only guaranteed to be defined on the interval $[0, 2/h]$. It is increasing and convex on $[0, 1/h]$, and for this reason we will subsequently ensure $0 \le q \le 1/h$ whenever we use $\Psi_{h,\tau}(q)$.

The evolution equations and their Bernstein transforms now take the following form.

Lemma 6.1. *Using the rescaled variables in (6.1), (5.7) becomes*

$$\frac{\langle f, \nu_{n+1}^h\rangle - \langle f, \nu_n^h\rangle}{\tau} = \sum_{k\ge 2} R_k^{h,\tau}(\rho_n^h) I_k(\nu_n^h, f) \quad \textit{for any bounded } f \in C(\mathbb{R}_+).$$

Here I_k is defined by (5.5), and

$$R_k^{h,\tau}(\rho_n^h) := \frac{(-\rho_n^h)^k}{k!} \Psi_{h,\tau}^{(k)}(\rho_n^h),$$

where $\rho_n^h = \langle 1, \nu_n^h\rangle$ represents the rescaled total number at time $n\tau$.

Further, the Bernstein transform of ν_n^h satisfies

$$\frac{\varphi_{n+1}^h(q) - \varphi_n^h(q)}{\tau} = -\Psi_{h,\tau}(\varphi_n^h(q)) \quad \textit{for all } n \ge 0,\ q \ge 0. \tag{6.3}$$

Proof. The proof is a direct computation using (5.7), Lemma 4.2, and (4.3). □

6.2. Convergence of critical branching mechanisms

The first step in studying continuum limits of ν_n^h is to study convergence of the branching mechanisms. We obtain necessary and sufficient conditions for such convergence in terms of a criterion that is closely tied to the continuity theorems relating Bernstein transforms and Lévy triples, which we develop in Appendix A.

For greater generality, we will study sequential limits where we also allow the measure $\hat{\pi}$ to vary. Let $(\hat{\pi}_k)$ be a sequence of probability measures on $\mathbb{N}_0$ which satisfies criticality condition (2.4), and let (h_k) and (τ_k) be positive sequences converging to 0. We introduce (rescaled) discrete branching mechanisms as in (4.4) and (6.2), with $\hat{\pi}_k$ replacing $\hat{\pi}$, by defining

$$\hat{\Psi}_k(q) := \sum_{j=2}^{\infty} ((1-q)^j - 1 + jq)\hat{\pi}_k(j), \qquad \check{\Psi}_k(q) := \frac{\hat{\Psi}_k(h_k q)}{\tau_k h_k}.$$

We next associate to each family-size distribution $\hat{\pi}_k$ a (Lévy) measure $\hat{\mu}_k$ on $\mathbb{R}^+$ given by

$$d\hat{\mu}_k(x) := \sum_{j\geq 2}(j-1)\hat{\pi}_k(j)\, d\delta_j(x).$$

This measure is rescaled according to

$$d\breve{\mu}_k(x) := \frac{1}{\tau_k}\, d\hat{\mu}_k\left(\frac{x}{h_k}\right) = \frac{1}{\tau_k}\sum_{j\geq 2}(j-1)\hat{\pi}_k(j)\, d\delta_{jh_k}(x). \tag{6.4}$$

In the next definition we single out a particular sense of convergence of these measures that will be important throughout the rest of this paper. This notion relates to convergence of Lévy triples and is revisited in Appendix A (see Remark A.1).

Definition 6.1. Given some finite measure κ on $[0,\infty]$, we say that the sequence $(\breve{\mu}_k)$ *Lévy-converges* to κ provided

$$(x\wedge 1)\, d\breve{\mu}_k(x) \to d\kappa(x) \quad \text{weak-}\star \text{ on } [0,\infty]. \tag{6.5}$$

Recall that a sequence of finite measures (κ_k) converges to κ weak-$\star$ on $[0,\infty]$ if, for every test function $g \in C([0,\infty])$, we have $\langle g,\kappa_k\rangle \to \langle g,\kappa\rangle$. We require test functions to be continuous at ∞ in order to capture any atom at ∞.

Proposition 6.1. *Given a sequence $(\hat{\pi}_k)$ satisfying criticality condition (2.4), and positive sequences (h_k) and (τ_k) converging to 0, let $(\breve{\mu}_k)$ and $(\breve{\Psi}_k)$ be as above.*

(i) *Suppose that $(\breve{\mu}_k)$ Lévy-converges to some finite measure κ on $[0,\infty]$ as $k\to\infty$. Then, for each $q\in[0,\infty)$ as $k\to\infty$, we have*

$$\breve{\Psi}_k(q) \to \Psi(q) := \frac{1}{2}\alpha_0 q^2 + \alpha_\infty q + \int_0^\infty \frac{e^{-qx}-1+qx}{x}\, d\mu(x), \tag{6.6}$$

where $(\alpha_0,\alpha_\infty,\mu)$ is the Lévy triple associated with κ by the relation

$$d\kappa(x) = \alpha_0\, d\delta_0 + \alpha_\infty\, d\delta_\infty + (x\wedge 1)\, d\mu(x). \tag{6.7}$$

Moreover, each derivative $\breve{\Psi}_k^{(m)}$ converges to $\Psi^{(m)}$, locally uniformly in $(0,\infty)$ for each $m\in\mathbb{N}$.

(ii) *Conversely, suppose $\Psi(q) = \lim_{k\to\infty}\breve{\Psi}_k(q)$ exists for each $q\in[0,\infty)$. Then $(\breve{\mu}_k)$ Lévy-converges to some finite measure κ on $[0,\infty]$, and Ψ is given by (6.6) and (6.7).*

Remark 6.1. The Lévy-convergence requirement (6.5) corresponds exactly to convergence of Lévy triples in a natural topology associated to subordinators. Explicitly, criterion (6.5) is equivalent to convergence of the Lévy triples $(0,0,\breve{\mu}_k)$ to the Lévy triple $(\alpha_0,\alpha_\infty,\mu)$, as described in Appendix A.

Remark 6.2. Recall that the expression for Ψ in (6.6) is the general form of a branching mechanism for a finite (sub)critical CSBP (see [25, Chapter 12] or Section 5). We show, in Section 8, that every such branching mechanism does arise as a sequential limit from discrete branching mechanisms of *critical* GW processes. A heuristic explanation as to why critical branching mechanisms might yield a subcritical branching mechanism in the limit is discussed in Remark 6.6, below.

https://doi.org/10.1017/apr.2018.23 Published online by Cambridge University Press

Proof of Proposition 6.1. (i) The proof follows in three steps.

Step 1. Define

$$d\kappa_k = (x \wedge 1)\, d\breve{\mu}_k. \tag{6.8}$$

Then $\kappa_k \to \kappa$ weak-$\star$ on $[0, \infty]$ by (6.5). Next fix $q \geq 0$ and compute

$$\Psi'(q) = \alpha_0 q + \alpha_\infty + \int_0^\infty (1 - e^{-qx})\, d\mu(x) = \langle f_0, \kappa \rangle, \tag{6.9}$$

$$\breve{\Psi}'_k(q) = \int_{[2h_k,\infty)} \frac{1 - (1 - h_k q)^{(x-h_k)/h_k}}{x - h_k} x\, d\breve{\mu}_k(x) = \langle f_{h_k}, \kappa_k \rangle, \tag{6.10}$$

where

$$f_0(x) = \frac{1 - e^{-qx}}{x}\left(\frac{x}{x \wedge 1}\right), \qquad f_h(x) = \frac{1 - (1 - hq)^{(x-h)/h}}{x - h}\left(\frac{x}{x \wedge 1}\right).$$

The second equality in (6.10) follows since κ_k is supported on $[2h_k, \infty)$.

Note that $\Psi(0) = \Psi_{h_k}(0) = 0$, and Ψ' and Ψ'_{h_k} are positive and increasing. Thus, the desired conclusion in (6.6) for fixed q will follow, provided we show that

$$\langle f_{h_k}, \kappa_k \rangle = \langle f_{h_k} - f_0, \kappa_k \rangle + \langle f_0, \kappa_k \rangle \to \langle f_0, \kappa \rangle. \tag{6.11}$$

Clearly, $\langle f_0, \kappa_k \rangle \to \langle f_0, \kappa \rangle$ since $f_0 \in C([0, \infty])$. We claim that

$$\lim_{h\to 0} \sup_{x>h} |f_h(x) - f_0(x)| = 0. \tag{6.12}$$

Since $\langle 1, \kappa_k \rangle \to \langle 1, \kappa \rangle$, this immediately implies (6.11).

Step 2. To complete the proof of (6.6), it only remains to prove (6.12). For this, we claim that, provided $\max(h, qh) < \frac{1}{2}$,

$$|f_h(x) - f_0(x)| \leq \begin{cases} 5e^{-qx} + \dfrac{2h}{x} & \text{for } x > 2, \\ 2(1 \vee x)hq^2 & \text{for } x > h. \end{cases} \tag{6.13}$$

Indeed, this estimate in the $x > 2$ case follows immediately from the bounds

$$0 < \frac{x}{x-h} - 1 = \frac{h}{x-h} < \frac{2h}{x}, \qquad \frac{x}{x-h}(1 - qh)^{(x-h)/h} \leq 2\frac{(e^{-qh})^{x/h}}{1 - qh} \leq 4e^{-qx}.$$

In the $x > h$ case, observe that since $z \mapsto e^z$ is contractive for $z < 0$, we have

$$\frac{|(1 - qh)^{(x-h)/h} - e^{-(x-h)q}|}{x - h} \leq \frac{1}{h}|\ln(1 - qh) + qh| = \frac{1}{h}\int_0^{qh} \frac{z}{1 - z}\, dz \leq hq^2,$$

since $qh < \frac{1}{2}$. Moreover, since $z \mapsto (1 - e^{-z})/z = \int_0^1 e^{-rz}\, dr$ is a decreasing contraction for $z > 0$,

$$0 < \frac{1 - e^{-(x-h)q}}{x - h} - \frac{1 - e^{-xq}}{x} \leq |(x - h)q - xq|q = hq^2.$$

By adding these last two bounds and using $x/(x \wedge 1) = 1 \vee x$, we infer (6.13).

Using the first estimate in (6.13) for $x > h^{-1/2}$, and the second estimate in (6.13) for $x \leq h^{-1/2}$, we obtain (6.12). This complete the proof of (6.6).

https://doi.org/10.1017/apr.2018.23 Published online by Cambridge University Press

Step 3. To prove the statement regarding local uniform convergence, let Ω be an open set with compact closure in the right half-plane $\{q \in \mathbb{C} \mid \operatorname{Re} q > 0\}$, and note that $|1 - hq| < 1$ for all $q \in \Omega$, for sufficiently small $h > 0$. When $x = mh$ for an integer $m \geq 2$, the function $q \mapsto f_h(x)$ is analytic and is clearly bounded on Ω uniformly for h small and for $x \geq 1$. A uniform bound holds for $x \leq 1$ as well due to the fact that in this case $\partial_q f_h(x) = (1-hq)^{m-2}$, whence $|\partial_q f_h(x)| \leq 1$.

Now, due to (6.10), the functions $(\breve{\Psi}'_k)_{k\geq N}$ are analytic and uniformly bounded on Ω. By Montel's theorem, this sequence converges uniformly on Ω, and by Cauchy's integral formula, all derivatives converge locally uniformly. This concludes the proof of (i).

(ii) Assume that $\Psi(q) = \lim_{k\to\infty} \breve{\Psi}_k(q)$ exists for each $q \in [0, \infty)$. First, considering some q fixed, note that, for every $x > h$,

$$f_h(x) \geq \frac{1 - \mathrm{e}^{-q(x-h)}}{x - h}(1 \vee x) \geq \frac{1 - \mathrm{e}^{-qx}}{x}(1 \vee x) = f_0(x).$$

Therefore, by (6.10), it follows that whenever k is so large that $2qh_k < 1$,

$$(\inf\, f_0)\langle 1, \kappa_k\rangle \leq \langle f_{h_k}, \kappa_k\rangle = \breve{\Psi}'_k(q) \leq \frac{\breve{\Psi}_k(2q)}{q}. \tag{6.14}$$

The last inequality holds since $\breve{\Psi}_k$ is convex and positive on $(q, 2q)$. Because $\inf f_0 > 0$ and $(\breve{\Psi}_k(2q))$ is bounded, it follows that $\sup_k\langle 1, \kappa_k\rangle < \infty$ and, hence, $\{\kappa_k\}$ is weak-$\star$ pre-compact.

Thus, any subsequence of (κ_k) has a further subsequence that converges weak-$\star$ on $[0, \infty]$. Let κ denote any such limit. By the proof of (i) above, we infer that, for any $q > 0$, as $k \to \infty$ along the appropriate subsequence,

$$\breve{\Psi}'_k(q) = \langle f_{h_k}, \kappa_k\rangle \to \Phi(q) := \langle f_0, \kappa\rangle.$$

By dominated convergence, we deduce that $\Psi(q) = \lim \int_0^q \breve{\Psi}'_k(r)\,\mathrm{d}r = \int_0^q \Phi(r)\,\mathrm{d}r$, whence we conclude Ψ is C^1 and $\Phi = \Psi'$.

Let $(\alpha_0, \alpha_\infty, \mu)$ be the Lévy triple associated with κ by (6.7). Then we see that (6.9) holds and Ψ' is the Bernstein transform of this triple. Since the Bernstein transform is bijective (see Theorem A.1), it follows that Ψ determines κ uniquely. As a consequence, the whole sequence (κ_k) must converge, meaning that (6.5) holds. □

6.3. Convergence of Bernstein transforms

Next we study the convergence of the rescaled size distributions ν_n^h for survivors, by the simple expedient of studying convergence of their Bernstein transforms φ_n^h, regarding (6.3) as a forward Euler difference approximation to the ordinary differential equation (5.3).

Proposition 6.2. *We make the same assumptions as in Proposition 6.1.*

(i) *Suppose that the Lévy-convergence in (6.5) holds for some finite measure κ on* $[0, \infty]$. *Then, as $k \to \infty$ and whenever $n\tau_k \to t$, we have*

$$\varphi_n^{h_k}(q) \to \varphi(q, t) \quad \textit{for every } q, t \in [0, \infty), \tag{6.15}$$

where φ is the unique solution of

$$\partial_t\varphi(q, t) = -\Psi(\varphi(q, t)), \qquad \varphi(q, 0) = q, \tag{6.16}$$

with Ψ *given by (6.6). Moreover,* $\varphi(\cdot, t)$ *is Bernstein for each* $t \geq 0$ *and has the form*

$$\varphi(q, t) = \beta_0(t)q + \beta_\infty + \int_{(0,\infty)} (1 - \mathrm{e}^{-qx}) \, \mathrm{d}\nu_t(x), \tag{6.17}$$

where $\beta_\infty = 0$*,* $\beta_0(t) \geq 0$*, and* $\int_{(0,\infty)} (x \wedge 1) \, \mathrm{d}\nu_t(x) < \infty$.

(ii) *Conversely, suppose that (6.15) holds, and* $\varphi(q_0, t_0) > 0$ *for some* $q_0, t_0 > 0$*. Then (6.5) holds for some finite measure* κ *on* $[0, \infty]$.

Proof. (i) The proof follows from providing a rather straightforward proof that the forward Euler difference scheme in (6.3) converges if (6.6) holds. Fix $q \in \mathbb{R}^+$ and $T > 0$. From (6.16), we see that

$$\varphi(q, t) = q - \int_0^t \Psi(\varphi(q, s)) \, \mathrm{d}s.$$

Since $s \mapsto \Psi(\varphi(q, s))$ is positive decreasing, $\varphi(q, s) \leq q$. Whenever $n\tau_k \leq T$, we have

$$\varphi(q, n\tau_k) = q - \tau_k \sum_{m=0}^{n-1} \Psi(\varphi(q, m\tau_k)) + R_1',$$

where

$$0 \leq R_1' \leq \tau_k \sum_{m=0}^{n-1} (\Psi(\varphi(q, m\tau_k)) - \Psi(\varphi(q, (m+1)\tau_k))) \leq \tau_k \Psi(q).$$

Let $\varphi_k(q, n\tau_k) := \varphi_n^{h_k}(q)$ and $\Psi_k := \Psi_{h_k,\tau_k}$. Summing (6.3) and noting that $\nu_0^h = \delta_h/h$, we have

$$\varphi_k(q, n\tau_k) = \frac{1 - \mathrm{e}^{-qh_k}}{h_k} - \tau_k \sum_{m=0}^{n-1} \Psi_k(\varphi_k(q, m\tau_k)) = q - \tau_k \sum_{m=0}^{n-1} \Psi(\varphi_k(q, m\tau_k)) + R_2', \tag{6.18}$$

where, since $n\tau_k \leq T$,

$$|R_2'| \leq M_k := q^2 h_k + T \sup_{[0,q]} |\Psi_k - \Psi|.$$

Consequently, since Ψ' is positive and increasing on $[0, q]$,

$$\begin{aligned} |(\varphi - \varphi_k)(q, n\tau_k)| &\leq |R_1'| + |R_2'| + \tau_k \sum_{m=0}^{n-1} |\Psi(\varphi(q, m\tau_k)) - \Psi(\varphi_k(q, m\tau_k))| \\ &\leq \tau_k \Psi(q) + M_k + \tau_k \Psi'(q) \sum_{m=0}^{n-1} |(\varphi - \varphi_k)(q, m\tau_k)|. \end{aligned}$$

Hence, by the discrete Gronwall inequality,

$$|(\varphi - \varphi_k)(q, n\tau_k)| \leq (\tau_k \Psi(q) + M_k)(1 + \tau_k \Psi'(q))^n \leq (\tau_k \Psi(q) + M_k) \, \mathrm{e}^{\Psi'(q) n\tau_k}.$$

Proposition 6.1 guarantees that $M_k \to 0$ as $k \to \infty$. Since $\varphi(q, t)$ has bounded derivative $|\partial_t \varphi| \leq \Psi(q)$, we may infer that whenever $n\tau_k \to t \in [0, T)$,

$$|\varphi_k(q, n\tau_k) - \varphi(q, t)| \to 0 \quad \text{as } k \to \infty.$$

https://doi.org/10.1017/apr.2018.23 Published online by Cambridge University Press

This completes the proof of the convergence of (φ_k). Since $\varphi_k(\cdot, n\tau_k)$ is Bernstein, and any pointwise limit of Bernstein functions is Bernstein (see [38, Corollary 3.7]), the limit $\varphi(\cdot, t)$ is Bernstein and has the form of (6.17) for some Lévy triple $(\beta_0, \beta_\infty, \nu_t)$. We must have $\beta_\infty = 0$, however, since $\varphi(q, t) \le q$ for all $q, t \ge 0$.

(ii) Choose $q, t > 0$ such that $\varphi(q, t) > 0$. From (6.18), we find that $\varphi_k(q, m\tau_k) \ge \varphi_k(q, n\tau_k)$ for $m \le n$. Consequently, for sufficiently large k with $n = \lfloor t/\tau_k \rfloor$, we have $n\tau_k > t/2$ and $\varphi_k(q, n\tau_k) > \varphi(q, t)/2$, and, from (6.18), we again find that

$$\frac{t}{2}\Psi_k\left(\frac{\varphi(q, t)}{2}\right) \le \tau_k \sum_{m=0}^{n-1} \Psi_k(\varphi_k(q, m\tau_k)) \le \frac{1 - \mathrm{e}^{-qh_k}}{h_k} \le q.$$

Hence, $\{\Psi_k(\varphi(q, t)/2)\}$ is bounded. Following the proof of Proposition 6.1 (specifically, using (6.14)), we see that the set of measures $\{\kappa_k\}$ is weak-$\star$ pre-compact. (Here κ_k is defined by (6.8).)

Thus, any subsequence of (κ_k) has a further subsequence that converges weak-$\star$ on $[0, \infty]$. Let κ denote any such limit. Then, by (i) above, we know (6.16) holds with Ψ given by taking the limit in (6.6) along the appropriate subsequence, as in the proof of Proposition 6.1(ii). Then φ is C^1 and Ψ is determined by φ by evaluating (6.16) at $t = 0$. As in the proof of Proposition 6.1, it follows that κ is determined by φ, hence the whole sequence (κ_k) converges. This means (6.5) holds, completing the proof. □

Since convergence of the Bernstein transforms is equivalent to convergence of the corresponding Lévy triples and weak-$\star$ convergence of the corresponding measures on $[0, \infty]$ by the continuity theorem (Theorem A.2), we obtain the following corollary.

Corollary 6.1. *Under the same assumptions as Proposition 6.2(i), it follows that, for every $t \ge 0$, as $k \to \infty$, and whenever $nh_k \to t$,*

$$(x \wedge 1)\,\mathrm{d}\nu_n^{h_k}(x) \to \beta_0(t)\,\mathrm{d}\delta_0(x) + (x \wedge 1)\,\mathrm{d}\nu_t(x) \quad \textit{weak-}\star \textit{ on } [0, \infty]. \tag{6.19}$$

Conversely, if this convergence holds then (6.5) holds for some finite measure κ on $[0, \infty]$.

Further, we have $\beta_0(t) = 0$ for all $t > 0$ if and only if, in (6.6), we have $\Psi'(\infty) = \infty$ or, equivalently,

$$a_0 > 0 \quad \textit{or} \quad \int_{(0,\infty)} \mathrm{d}\mu(x) = \infty. \tag{6.20}$$

In the $\Psi'(\infty) < \infty$ case, $\beta_0(t) = \exp(-\Psi'(\infty)t)$.

Moreover, we have $\varphi(\infty, t) = \langle 1, \nu_t \rangle < \infty$ if and only if Grey's condition holds:

$$\int_{[1,\infty)} \frac{\mathrm{d}u}{\Psi(u)} < \infty. \tag{6.21}$$

In this case, the limiting family of finite measures $(\nu_t)_{t \ge 0}$ is a weak solution of the generalized damped Smoluchowski equation

$$\partial_t \langle f, \bar{\nu}_t \rangle = -\Psi'(0^+)\langle f, \bar{\nu}_t \rangle + \sum_{k \ge 2} R_k I_k(\bar{\nu}_t, f) \tag{6.22}$$

for all test functions $f \in C([0, \infty])$. Here R_k and I_k are as in (5.5) and (5.6), respectively.

https://doi.org/10.1017/apr.2018.23 Published online by Cambridge University Press

Remark 6.3. In the critical case, $\Psi'(0^+) = 0$ and (6.22) is precisely the generalized Smoluchowski equation (5.4) encountered before.

Remark 6.4. The convergence property (6.19) explains precisely how the Lévy jump measure of a (sub)critical CSBP X arises as a limit of scaled nth generation descendant distributions. This was one of our main motivations for developing this treatment of continuum limits of GW processes.

Remark 6.5. For the case when Grey's condition (6.21) does not hold, Bertoin and Le Gall proposed an interesting generalization of Smoluchowski's equation in terms of the sum of locations of atoms of Poisson random measures; see [13, Equation (26)].

Proof of Corollary 6.1. The proof follows in four steps.

Step 1. The weak convergence in (6.19) and the converse follow immediately from Proposition 6.1 and the continuity theorem for Bernstein transforms (Theorem A.2).

Step 2. Next, note that by (6.17) and since $(1 - \mathrm{e}^{-qx})/q \le x \wedge (1/q) \to 0$ as $q \to \infty$,

$$\beta_0(t) = \lim_{q\to\infty} \frac{\varphi(q,t)}{q}. \tag{6.23}$$

We now claim that $\beta_0(t) = 0$ for all $t > 0$ if and only if $\lim_{q\to\infty} \Psi'(q) = \infty$. Indeed, if $\beta_0(t) \ge \varepsilon > 0$ for some $t > 0$ then, since $\varphi(q,t)$ is concave in q and decreasing in t, necessarily $\varphi(q,t) \ge \varepsilon q$ and

$$q \ge \int_0^t \Psi(\varphi(q,s))\,\mathrm{d}s \ge t\Psi(\varepsilon q) \quad \text{for all } q > 0.$$

Since Ψ is convex, it follows that $\Psi'(q) \le 1/(\varepsilon t)$ for all q. Conversely, if $\Psi'(q) \le M$ for all q then $\Psi(q) \le Mq$ and $\varphi(q,t) \ge q\mathrm{e}^{-Mt}$ by (6.16).

Now, the condition $\lim_{q\to\infty} \Psi'(q) = \infty$ is easily seen to be equivalent to (6.20). This establishes both as necessary and sufficient to have $\beta_0(t) \equiv 0$. In the $\Psi'(\infty) < \infty$ case, due to (6.16) and (6.23), we find that

$$\beta_0(t) - \beta_0(s) = -\lim_{q\to\infty} \frac{1}{q} \int_s^t \Psi(\varphi(q,r))\,\mathrm{d}r = -\int_s^t \Psi'(\infty)\beta_0(r)\,\mathrm{d}r,$$

whence $\beta(t) = \exp(-\Psi'(\infty)t)$ since $\beta_0(0) = 1$.

Step 3. The fact that Grey's condition is necessary and sufficient for $\langle 1, \nu_t\rangle < \infty$ is well known (see, for example, the proof of [25, Theorem 12.5]) and is easily deduced by separating variables in (6.16) and integrating.

Step 4. Finally, we show that (ν_t) satisfies (6.22). When Ψ is critical (i.e. $\Psi'(0^+) = 0$), a proof can be found in [13, Proposition 3] or [22, Theorem 3.2]. In the general case, the proof closely resembles that of [22, Theorem 3.2] and we sketch the details here.

The main idea is to establish (6.22) for test functions of the form $f_q(x) := 1 - \mathrm{e}^{-qx}$. Indeed, observe that

$$\begin{aligned}
\rho_t^k I_k(\bar\nu, f_q) &= \int_{(0,\infty)^k} \biggl[f_q\biggl(\sum_{i=1}^k x_i\biggr) - \sum_{i=1}^k f_q(x_i)\biggr] \mathrm{d}\bar\nu_t^k \\
&= \int_{(0,\infty)^k} \biggl[1 - \prod_{i=1}^k \mathrm{e}^{-qx_i} - \sum_{i=1}^k (1 - \mathrm{e}^{-qx_i})\biggr] \mathrm{d}\bar\nu_t^k \\
&= \rho_t^k - (\rho_t - \varphi(q,t))^k - k\rho_t^{k-1}\varphi(q,t).
\end{aligned}$$

https://doi.org/10.1017/apr.2018.23 Published online by Cambridge University Press

Consequently,

$$\begin{aligned}\sum_{k\geq 2} R_k I_k(\bar{\nu}, f_q) &= \sum_{k\geq 2} \frac{(-1)^k \Psi^{(k)}(\rho_t)}{k!}[\rho_t^k - (\rho_t - \varphi(q,t))^k - k\rho_t^{k-1}\varphi(q,t)] \\ &= \sum_{k\geq 0} \frac{(-1)^k \Psi^{(k)}(\rho_t)}{k!}[\rho_t^k - (\rho_t - \varphi(q,t))^k - k\rho_t^{k-1}\varphi(q,t)] \\ &= \Psi(0) - \Psi(\varphi(q,t)) + \varphi(q,t)\Psi'(0) \\ &= \varphi(q,t)\Psi'(0) + \partial_t\langle\bar{\nu}, f_q\rangle,\end{aligned}$$

where the second equality holds since the terms for $k = 0, 1$ are 0, the third equality follows from the Taylor expansion of Ψ about ρ_t, and the last equality follows from (6.16) and the fact that $\Psi(0) = 0$.

This proves that (6.22) is satisfied for test functions of the form f_q above. The general case follows from an approximation argument, the details of which are the same as in the proof of [22, Theorem 3.2]. □

6.4. Convergence of GW processes

Finally, we conclude this section with convergence results for the rescaled GW processes. We rescale the population by a factor of h_k and the reproduction time by a factor of τ_k. Explicitly, the rescaled GW process $Y^{(k)}$ is defined by

$$Y_t^{(k)}(x) = h_k X_{\lfloor t/\tau_k \rfloor, k}\left(\left\lfloor \frac{x}{h_k} \right\rfloor\right). \tag{6.24}$$

We recall that the argument of the processes refers to the initial population and is usually suppressed.

Proposition 6.3. *We make the same assumptions as in Proposition 6.1.*

(i) *Suppose there exists a finite measure κ on $[0, \infty]$ such that $(\breve{\mu}_k)$ Lévy-converges (as in Definition 6.1) to κ. Then the finite-dimensional distributions of the rescaled GW processes $Y^{(k)}$ converge to those of a finite (sub)critical CSBP Z. Further, the Bernstein transforms of the $Y^{(k)}$ converge pointwise to the Laplace exponent of Z. That is, if*

$$\varphi_k(q,t) := \frac{1}{h_k}\mathbb{E}[1 - \exp(-qY_t^{(k)}(h_k))] \quad \textit{and} \quad \varphi(q,t) := -\ln \mathbb{E}[\exp(-qZ_t(1))],$$

then

$$\lim_{k\to\infty} \varphi_k(q,t) = \varphi(q,t) \quad \textit{for every } q \in \mathbb{R}^+ \textit{ and } t \geq 0.$$

(ii) *Conversely, if the one-dimensional distributions of $Y^{(k)}$ converge to those of a nonnegative process Z such that $\mathbb{P}\{Z_{t_0}(x_0) > 0\} > 0$ for some $x_0, t_0 > 0$, then there exists a finite measure κ on $[0, \infty]$ such that $(\breve{\mu}_k)$ Lévy-converges to κ. Consequently, Z may be chosen to be a finite (sub)critical CSBP.*

Remark 6.6. Before presenting the proof, we momentarily pause to remark on criticality. Each process $Y^{(k)}$ is a critical GW process, however, the limit need not be critical. Indeed, the branching mechanism of the limit is given by (6.6) with $\alpha_\infty \geq 0$. This means that the limiting process Z can either be *critical* or *subcritical*. The subcritical situation ($\alpha_\infty > 0$) arises precisely when the rescaled measures $\breve{\mu}_k$ Lévy-converge to a measure that has an atom at ∞. Physically, this corresponds to the situation when the rescaled descendant distributions $\breve{\mu}_k$ have a negligible fraction of large families that contain a non-negligible fraction of the population. In the continuum limit, this fraction of the population ends up in families of 'infinite' size, and the total population observed in families of finite size decays in time at exponential rate a_∞.

Proof of Proposition 6.3. First assume that (6.5) holds. By Proposition 6.2, we know that φ_k converges to a function φ which satisfies (6.16). We know from [22] (see also [13]) that the solution of (6.16) is the Laplace exponent of a (sub)critical CSBP Z. To conclude the proof we only need to show that the finite-dimensional distributions of $Y^{(k)}$ converge to that of Z. Let $x > 0$ be the initial population and fix $t > 0$. Define $N_k = \lfloor x/h_k \rfloor$ and observe that

$$\begin{aligned}\mathbb{E}(\exp(-qY_t^{(k)}(x))) &= \mathbb{E}(\exp(-qY_t^{(k)}(N_k h_k))) \\ &= [\mathbb{E}(\exp(-qY_t^{(k)}(h_k)))]^{N_k} \\ &= [1 - h_k\varphi_k(q,t)]^{N_k} \\ &\to \exp(-\varphi(q,t)x) \\ &= \mathbb{E}(\exp(-qZ_t(x))), \qquad k \to \infty.\end{aligned}$$

From this we see that the Laplace transforms of $Y_t^{(k)}$ converge to the Laplace transform of Z, proving convergence of one-dimensional distributions. The convergence of higher-dimensional distributions can now be obtained as in [29, p. 280].

For the converse, assume that the one-dimensional distributions of $Y^{(k)}$ converge to that of a process Z with $\mathbb{P}\{Z_{t_0} > 0\} > 0$. Let $x > 0$, $N_k = \lfloor x/h_k \rfloor$, and observe that convergence of one-dimensional distributions implies

$$\mathbb{E}(\exp(-qY_{t_0}^{(k)}(x))) \to \exp(-\varphi(q,t_0)x) = \mathbb{E}(\exp(-qZ_{t_0}(x))), \qquad k \to \infty.$$

Thus,

$$\lim_{k\to\infty} [1 - h_k\varphi_k(q,t_0)]^{N_k}$$

exists and is equal to $\exp(-\varphi(q,t_0)x)$ for some function φ. Clearly, $\varphi(q,t_0)x$ must be the Laplace exponent of $Z_{t_0}(x)$ and, further,

$$\lim_{k\to\infty} \varphi_k(q,t_0) = \lim_{k\to\infty} \frac{1 - \exp(-\varphi(q,t_0)x/N_k)}{h_k} = \varphi(q,t_0).$$

By our assumption on Z, we know that $\varphi(q,t_0) > 0$ and we can apply Proposition 6.2. This will guarantee that (6.5) holds. Using part (i), we infer that there is a finite (sub)critical CSBP with the same one-dimensional distributions as Z. This completes the proof. □

As an immediate corollary, we prove a special case of a result of [13] which establishes that the Lévy jump measure of certain CSBPs satisfies the Smoluchowski equation.

Corollary 6.2. *Let Z be a finite (sub)critical CSBP whose branching mechanism satisfies Grey's condition (6.21), and ν_t be the Lévy jump measure associated with Z_t. Then ν is a weak solution of the damped Smoluchowski equations (6.22).*

Proof. We first note that there exists a sequence of processes $Y^{(k)}$ of the form (6.24) such that the finite-dimensional distributions of $Y^{(k)}$ converge to Z. While the existence of such a sequence is well known (see [19] and [31]), we will prove a much stronger result; see Theorem 8.1 below. By Proposition 6.3, we know that the Bernstein transforms of $Y^{(k)}$ converge pointwise to the Laplace exponent of Z. Consequently, by Corollary 6.1, the convergence (6.19) holds, where ν_t is the Lévy jump measure of Z_t and, further, ν satisfies (6.22), as desired. □

7. Comparison to general convergence criteria

Necessary and sufficient criteria for convergence of scaling limits of GW processes in general were developed by Grimvall [19] and more recently by Bansaye and Simatos [4]. Grimvall's original criteria involved convolution powers of family-size distributions, shifted and scaled. The (equivalent) convergence criteria of Bansaye and Simatos [4], obtained in connection with their more general investigation of branching in time-varying environments, are simpler and only involve convergence of moments and tails of the family-size distribution. We present these criteria here and show that for critical GW processes the Bansaye–Simatos criteria are equivalent to the Lévy-convergence criterion in Proposition 6.3.

7.1. The Grimvall and Bansaye–Simatos convergence criteria

We begin by stating Grimvall's result using notation consistent with ours.

Theorem 7.1. (Grimvall [19, Theorems 3.1 and 3.3].) *Let (h_k) and (τ_k) be positive sequences converging to 0 such that $1/(h_k\tau_k) \in \mathbb{N}$ for all k. Let (η_k) be a sequence of probability measures of the form*

$$\eta_k = \sum_{j\geq -1} \hat{\pi}_k(j+1)\delta_{jh_k}.$$

(i) *If $\eta_k^{*1/(h_k\tau_k)}$ converges weakly to a probability measure η then the finite-dimensional distributions of the scaled processes $Y^{(k)}$ defined in (6.24) converge to those of a (possibly infinite) CSBP with an absorbing state $+\infty$.*

(ii) *Conversely, if, for every $t \in [0,1]$, the random variables $Y_t^{(k)}$ converge weakly to a random variable Y_t, with $\mathbb{P}\{Y_1 > 0\} > 0$, then $\eta_k^{*1/(h_k\tau_k)}$ converges weakly to a probability measure η.*

Remark 7.1. The reason the shifted measures η_n are natural in this context is because they determine the law of the parent process of the Lamperti transform. At the discrete level before scaling, this can be described as follows. Let ξ be a random variable with distribution $\hat{\pi}$, representing the number of descendants of a single individual. Define the *parent process* $\bar{X}$ to be a random walk obtained by summing i.i.d. copies of $\xi - 1$, and a time change Θ defined recursively by

$$\Theta_0 = 0, \qquad \Theta_{n+1} := \Theta_n + x + \bar{X}_{\Theta_n}.$$

The Lamperti transformation defines a process X by

$$X_n := x + \bar{X}_{\Theta_n}.$$

Observe that

$$X_{n+1} - X_n = \bar{X}_{\Theta_{n+1}} - \bar{X}_{\Theta_n}$$

and, hence, the increment $X_{n+1} - X_n$ has the same distribution as the sum of X_n independent copies of $\xi - 1$ that are each independent of X_n. Equivalently, X_{n+1} has the same law as the sum of X_n i.i.d. copies of ξ, that are each independent of X_n. This implies X_n is a GW process

https://doi.org/10.1017/apr.2018.23 Published online by Cambridge University Press

with initial population x and descendant distribution $\hat{\pi}$. (This is the discrete analog of the Lamperti transform described in Section 5.)

Grimvall's convergence criterion stated above implies that the sequence of the parent processes $\bar{X}^{(k)}$ generated by $\hat{\pi}_k$ converge weakly, after scaling, to a Lévy process. The limiting parent process will be the parent process of the CSBP as given by the Lamperti transform.

In their investigation of scaling limits for GW processes in varying environments, Bansaye and Simatos [4] identified the following simplified criteria as equivalent to Grimvall's for the convergences in Theorem 7.1. In the present situation, the criteria of Bansaye and Simatos (stated as Assumption A.1 in Appendix A of [4]) may be stated as follows.

Proposition 7.1. (Bansaye and Simatos [4].) *The finite-dimensional distributions of the scaled processes $Y^{(k)}$ defined in (6.24) converge to those of a (possibly infinite) CSBP with an absorbing state $+\infty$ if and only if there exists $\hat{a} \in \mathbb{R}$, $\hat{b} \geq 0$, and a positive σ-finite measure F on $(0, \infty)$ such that, as $k \to \infty$,*

$$\int_{\mathbb{R}} \frac{x}{1+x^2} \frac{\mathrm{d}\eta_k(x)}{h_k \tau_k} \to \hat{a}, \qquad \int_{\mathbb{R}} \frac{x^2}{1+x^2} \frac{\mathrm{d}\eta_k(x)}{h_k \tau_k} \to \hat{b}, \tag{7.1}$$

and

$$\int_{[x,\infty)} \frac{\mathrm{d}\eta_k(x)}{h_k \tau_k} \to F([x, \infty)) \quad \textit{for a.e. } x > 0. \tag{7.2}$$

We remark that one can directly prove the equivalence of the criteria (7.1) and (7.2), and Grimvall's convergence criterion in Theorem 7.1(i) using the classical theory of infinite divisibility and canonical measures from [16].

7.2. Equivalence to Lévy-convergence in the critical case

The main result of this section establishes that under the criticality assumption (2.4), conditions (7.1) and (7.2) are equivalent to the Lévy-convergence (Definition 6.1) of the rescaled Lévy measures $\breve{\mu}_k$ determined from family-size distributions $\hat{\pi}_k$ as in (6.4). What makes this a curious point is that Lévy-convergence occurs only on the compactified positive half-line $[0, \infty]$, while the measures η_k also have support on the negative half-line.

Proposition 7.2. *Let $(\hat{\pi}_k)$ be a sequence of measures satisfying criticality assumption (2.4), (h_k) and (τ_k) be two positive sequences that converge to* 0*, and let $\breve{\mu}_k$ be defined by (6.4). Then (7.1) and (7.2) are equivalent to the existence of a finite measure κ on $[0, \infty]$ such that $(\breve{\mu}_k)$ Lévy-converges to κ.*

Proof. The proof follows in three steps.

Step 1. We note that by criticality and (6.4),

$$\frac{\hat{\pi}_k(0)}{\tau_k} = \frac{1}{\tau_k} \sum_{j \geq 2} (j-1) \hat{\pi}_k(j) = \int_{(0,\infty)} \mathrm{d}\breve{\mu}_k(x).$$

Recalling the definition of η_k in Theorem 7.1, we compute that

$$\int_{\mathbb{R}} x f(x) \frac{\mathrm{d}\eta_k(x)}{h_k \tau_k} = \sum_{j \geq 0} f(j h_k - h_k)(j-1) \frac{\hat{\pi}_k(j)}{\tau_k} = \int_{(0,\infty)} (f(x - h_k) - f(-h_k))\, \mathrm{d}\breve{\mu}_k(x)$$

whenever $x \mapsto x f(x)$ is bounded and continuous on $\mathbb{R}$.

https://doi.org/10.1017/apr.2018.23 Published online by Cambridge University Press

Step 2. Suppose that $(\breve{\mu}_k)$ Lévy-converges to a finite measure κ on $[0,\infty]$, and associate a Lévy triple $(\alpha_0,\alpha_\infty,\mu)$ with κ using (6.7). Then taking xf to approximate the characteristic function of $[z,\infty)$, we find that, for a.e. $z>0$,

$$\int_{[z,\infty)} \frac{\mathrm{d}\eta_k(x)}{h_k\tau_k} \to F([z,\infty)) = \int_{[z,\infty)} \frac{\mathrm{d}\mu(x)}{x}. \tag{7.3}$$

Further, taking f to be $x^p/(1+x^2)$ for $p=0,1$ we find that, as $k\to\infty$,

$$g_k(x) = \frac{f(x-h_k)-f(-h_k)}{x\wedge 1} = \frac{1}{x\wedge 1}\int_0^x f'(z-h_k)\,\mathrm{d}z \to \frac{f(x)-f(0)}{x\wedge 1}$$

uniformly on $[0,\infty]$. Hence, since the measures $(x\wedge 1)\,\mathrm{d}\breve{\mu}_k(x)$ are uniformly bounded and converge weak-$\star$ to κ on $[0,\infty]$, we have

$$\hat{a}_k := \int_{\mathbb{R}} \frac{x}{1+x^2}\frac{\mathrm{d}\eta_k(x)}{h_k\tau_k} \to \hat{a} = -\alpha_\infty - \int_{(0,\infty)} \frac{x^2}{1+x^2}\,\mathrm{d}\mu(x), \tag{7.4}$$

$$\hat{b}_k := \int_{\mathbb{R}} \frac{x^2}{1+x^2}\frac{\mathrm{d}\eta_k(x)}{h_k\tau_k} \to \hat{b} = \alpha_0 + \int_{(0,\infty)} \frac{x}{1+x^2}\,\mathrm{d}\mu(x). \tag{7.5}$$

Thus, both conditions (7.1) and (7.2) hold.

Step 3. Conversely, suppose that (7.1) and (7.2) hold. Taking $f(x)=(x-1)/(1+x^2)$ and supposing $h_k\in(0,1)$, we find that $g_k(x)\geq \frac{1}{2}$ for all x in the interval $[2h_k,\infty)$ containing the support of $\breve{\mu}_k$, and that, as $k\to\infty$,

$$\int_{(0,\infty)} g_k(x)(x\wedge 1)\,\mathrm{d}\breve{\mu}_k(x) = \hat{b}_k - \hat{a}_k \to \hat{b}-\hat{a}.$$

It follows that the measures $(x\wedge 1)\,\mathrm{d}\breve{\mu}_k(x)$ are uniformly bounded, hence, weak-$\star$ compact on $[0,\infty]$. For any subsequential limit κ on $[0,\infty]$ with associated Lévy triple $(\alpha_0,\alpha_\infty,\mu)$, relations (7.3)–(7.5) hold, and these uniquely determine the triple. Hence, the whole sequence $(\breve{\mu}_k)$ Lévy-converges. □

8. Universal GW family-size distributions

The following result is motivated by the existence of universal eternal solutions of Smoluchowski's coagulation equations [33], which, in turn, was inspired by Feller's account of Doeblin's universal laws in classical probability theory; see [16, Section XVII.9].

Theorem 8.1. *There exists a GW process X with family-size distribution $\hat{\pi}:\mathbb{N}\to[0,\infty)$ and sequences (h_n), $(\tau_n)\to 0$ with the following property. For any (sub)critical branching mechanism Ψ taking the form of (6.6), there is a subsequence along which the rescaled processes $Y^{(k)}$ (defined by rescaling X as in (6.24)) converge to a CSBP with branching mechanism Ψ.*

Moreover, the set of such family-size distributions $\hat{\pi}$ is dense in the set of all probability measures on $\mathbb{N}_0$ with the weak-$\star$ topology.

Much of the technical basis that we need to prove this result can be inferred directly from [33, Section 7]. The terminology used in [33], however, is substantially different from our terminology, which is based on [38] and Appendix A. Thus, for the convenience of the reader we provide a self-contained treatment here.

https://doi.org/10.1017/apr.2018.23 Published online by Cambridge University Press

We begin by constructing a 'universal' Lévy triple, in the sense that any other Lévy triple can be obtained as a suitable scaling limit. In order to make this precise, we need to define the correct notions of convergence and scaling of Lévy triples.

The correct notion of convergence of Lévy triples is a generalization of Lévy-convergence as introduced earlier in Definition 6.1. Namely, to each Lévy triple $\lambda := (\alpha_0, \alpha_\infty, \mu)$, we associate κ, a finite measure on $[0, \infty]$, by

$$\mathrm{d}\kappa(x) = a_0 \,\mathrm{d}\delta_0(x) + a_\infty \,\mathrm{d}\delta_\infty(x) + (x \wedge 1)\,\mathrm{d}\mu(x). \tag{8.1}$$

Now convergence of Lévy triples is defined using weak-$\star$ convergence of the associated κ-measures.

Definition 8.1. Let (λ_k) be a sequence of Lévy triples and κ_k the associated measures defined as in (8.1) above. We say that (λ_k) converges if the sequence (κ_k) converges weak-$\star$ in the space of finite measures on $[0, \infty]$.

A more detailed account of this can be found in Appendix A. Appendix A also contains variants of certain classical continuity theorems for Laplace transforms that are used throughout this section but are not widely known.

Next we define the scaling of Lévy triples as follows. Given a Lévy triple $\lambda = (\alpha_0, \alpha_\infty, \mu)$ and $b, c > 0$, define the rescaled Lévy triple $\lambda^{b,c}$ by

$$\lambda^{b,c} := (\alpha_0^{b,c}, \alpha_\infty^{b,c}, \mu^{b,c}),$$

where

$$\alpha_0^{b,c} = cb^{-1}\alpha_0, \qquad \alpha_\infty^{b,c} = c\,\alpha_\infty, \qquad \mathrm{d}\mu^{b,c}(x) = c\,\mathrm{d}\mu(bx). \tag{8.2}$$

In Section 9 we will see that this naturally corresponds to a dilational scaling of CSBPs. With this, we can define the aforementioned notion of universality.

Definition 8.2. Let $\lambda_\star$ be a Lévy triple, and (b_k) and (c_k) be two sequences that converge to ∞. We say that $\lambda_\star$ is a *universal Lévy triple* with sequences (b_k) and (c_k) if, for any Lévy triple λ, we have

$$\lambda_\star^{b_k,c_k} \to \lambda \quad \text{as } k \to \infty \text{ along some subsequence.} \tag{8.3}$$

In our next lemma we show that this notion of universality is completely determined by the tail of $\lambda_\star$.

Lemma 8.1. *Let $\lambda_\star = (0, 0, \mu_\star)$ be a universal Lévy triple with sequences (b_k) and (c_k). Let $\alpha_0 \geq 0$, $R > 0$, μ be any Lévy measure, and define the Lévy measure $\nu_\star$ by*

$$\nu_\star(A) = \mu(A \cap (0, R]) + \mu_\star(A \cap (R, \infty)).$$

Then the Lévy triple $(\alpha_0, 0, \nu_\star)$ is also universal with the sequences (b_k) and (c_k).

Remark 8.1. Lemma 8.1 still holds if finitely many terms of the sequences (b_k) and (c_k) are arbitrarily changed.

To prove Lemma 8.1 it helps to introduce left and right distribution functions as follows.

Definition 8.3. Given a Lévy triple λ, we define the pair of left and right distribution functions associated with λ to be (κ_L, κ_R) defined by

$$\kappa_L(x) := \alpha_0 + \int_{(0,x]} z\,\mathrm{d}\mu(z), \qquad \kappa_R(x) := \alpha_\infty + \int_{(x,\infty)} \mathrm{d}\mu.$$

https://doi.org/10.1017/apr.2018.23 Published online by Cambridge University Press

The reason we introduce these functions is because of a variant of a classical continuity theorem: pointwise convergence (almost everywhere or at points of continuity) of the pair of left and right distribution functions is equivalent to convergence of the associated Lévy triples. Since the proof in this form is not readily available in the literature, we provide a proof in Appendix A (Theorem A.3). We can now prove Lemma 8.1.

Proof of Lemma 8.1. Let λ be any Lévy triple and choose a subsequence for which the convergence in (8.3) holds. We claim that along this subsequence, we must have $(c_k/b_k) \to 0$. Once this is established, the lemma immediately follows from the fact that

$$(c_k b_k^{-1}\alpha_0, 0, |\mu_\star^{b_k,c_k} - \nu_\star^{b_k,c_k}|) \to 0 \quad \text{in the topology of Lévy triples.}$$

To show that $(b_k/c_k) \to 0$, let $(\kappa_{\star\mathrm{L}}, \kappa_{\star\mathrm{R}})$ be the pair of left and right distribution functions associated with $\lambda_\star$. Under scaling, note that

$$\kappa_{\star\mathrm{L}}^{b_k,c_k}(x) = \frac{c_k}{b_k}\kappa_{\star\mathrm{L}}(b_k x) \quad \text{and} \quad \kappa_{\star\mathrm{R}}^{b_k,c_k}(x) = c_k\kappa_{\star\mathrm{R}}(b_k x). \tag{8.4}$$

Now, by the monotonicity of $\kappa_{\star\mathrm{L}}$ and universality of $\lambda_\star$, it follows that $\kappa_{\star\mathrm{L}}(y) \to \infty$ as $y \to \infty$. Moreover, (8.3) and the continuity theorem (Theorem A.3) imply that $\kappa_{\star\mathrm{L}}^{b_k,c_k}(x)$ converges (along the chosen subsequence) to a finite limit for some x. This forces $(c_k/b_k) \to 0$ along the chosen subsequence, completing the proof. □

The main idea behind the proof of Theorem 8.1 is the existence of many universal Lévy triples. This is our next result.

Proposition 8.1. *There exist sequences* (b_k) *and* (c_k) *such that the set of Lévy triples of the form* $(0, 0, \mu_\star)$*, which are universal with respect to* (b_k) *and* (c_k)*, is dense in the space of all Lévy triples.*

We postpone the proof of Proposition 8.1 until the end of this section.

Proof of Theorem 8.1. Using Proposition 8.1, we choose a Lévy triple $\lambda_\star = (0, 0, \mu_\star)$ that is universal with sequences (b_k) and (c_k). Using Lemma 8.1, we can, without loss of generality, assume that

$$\int_{(0,\infty)} \mathrm{d}\mu_\star(x) < 1.$$

Let $\tau_k = c_k^{-1}$ and $h_k = b_k^{-1}$, define a family size distribution $\hat{\pi}(j)$, $j \geq 0$, by

$$\hat{\pi}(0) = 1 - \int_{\mathbb{R}^+} \mathrm{d}\mu_\star(x), \qquad \hat{\pi}(j) = \frac{1}{j-1}\int_{(j-2,j-1]} \mathrm{d}\mu_\star(x), \quad j \geq 2,$$

and set $\hat{\pi}(1)$ so that $\sum_{j\geq 0}\hat{\pi}(j) = 1$. Define the coarse-grained Lévy measure $\hat{\mu}$ by

$$\mathrm{d}\hat{\mu}(x) = \sum_{j\geq 2}(j-1)\hat{\pi}(j)\,\mathrm{d}\delta_j(x) = \sum_{j\geq 2}\biggl(\int_{(j-2,j-1]}\mathrm{d}\mu_\star(x)\biggr)\mathrm{d}\delta_j(x).$$

We claim that the Lévy triple $\hat{\lambda} := (0, 0, \hat{\mu})$ is also universal. Once the universality of $\hat{\lambda}$ is established, the subsequent convergence of $Y^{(k)}$ follows immediately from Proposition 6.3.

To prove the universality of $\hat{\lambda}$, let $\lambda = (\alpha_0, \alpha_\infty, \mu)$ be an arbitrary Lévy triple. By the universality of $\lambda_\star$, we have

$$\lambda_\star^{b_k,c_k} \to \lambda \quad \text{as } k \to \infty \text{ along some subsequence.}$$

For brevity, we use k to index the above subsequence.

https://doi.org/10.1017/apr.2018.23 Published online by Cambridge University Press

Let $(\kappa_{\mathrm{L}}, \kappa_{\mathrm{R}})$, $(\kappa_{\star\mathrm{L}}, \kappa_{\star\mathrm{R}})$, and $(\kappa_{\star\mathrm{L},k}, \kappa_{\star\mathrm{R},k})$ be the pairs of left and right distribution functions associated to the Lévy triples λ, $\lambda_\star$, and $\lambda_\star^{b_k,c_k}$, respectively. Since $\lambda_\star^{b_k,c_k} \to \lambda$, by the choice of our subsequence, we must have

$$\frac{h_k}{\tau_k}\kappa_{\star\mathrm{L}}\left(\frac{x}{h_k}\right) \to \kappa_{\mathrm{L}}(x) \quad \text{and} \quad \frac{1}{\tau_k}\kappa_{\star\mathrm{R}}\left(\frac{x}{h_k}\right) \to \kappa_{\mathrm{R}}(x)$$

at all points of continuity along the same subsequence. We claim that, along the same subsequence and at the same points, we must have

$$\frac{h_k}{\tau_k}\hat{\kappa}_{\mathrm{L}}\left(\frac{x}{h_k}\right) \to \kappa_{\mathrm{L}}(x) \quad \text{and} \quad \frac{1}{\tau_k}\hat{\kappa}_{\mathrm{R}}\left(\frac{x}{h_k}\right) \to \kappa_{\mathrm{R}}(x), \tag{8.5}$$

where $(\hat{\kappa}_{\mathrm{L}}, \hat{\kappa}_{\mathrm{R}})$ are the pair of left and right distribution functions associated to $\hat{\lambda}$.

To see this, observe that, for any $z \in (2, \infty)$,

$$\int_{(z,\infty)} \mathrm{d}\mu_\star(x) \le \sum_{j>z} \int_{(j-2,\, j-1]} \mathrm{d}\mu_\star(x) \le \int_{(z-2,\infty)} \mathrm{d}\mu_\star(x),$$

which means that

$$\kappa_{\star\mathrm{R}}(z) \le \hat{\kappa}_{\mathrm{R}}(z) \le \kappa_{\star\mathrm{R}}(z-2).$$

Then for any two points of continuity $0 < x_- < x$ for κ_{R}, whenever h_k is small enough we have $h_k^{-1}x_- < h_k^{-1}x - 2 < h_k^{-1}x$, whence

$$\kappa_{\mathrm{R}}(x) \le \liminf \frac{1}{\tau_k}\hat{\kappa}_{\mathrm{R}}\left(\frac{x}{h_k}\right) \le \limsup \frac{1}{\tau_k}\hat{\kappa}_{\mathrm{R}}\left(\frac{x}{h_k}\right) \le \kappa_{\mathrm{R}}(x_-).$$

Thus, the second limit in (8.5) follows.

To establish the first limit, observe that, for any $z > 2$,

$$\int_{(0,z-1]} x \,\mathrm{d}\mu_\star(x) \le \sum_{2\le j\le z} j \int_{(j-2,\, j-1]} \mathrm{d}\mu_\star(x) \le \int_{(0,z]} (x+2)\,\mathrm{d}\mu_\star(x),$$

and this entails

$$\kappa_{\star\mathrm{L}}(z-1) \le \hat{\kappa}_{\mathrm{L}}(z) \le \kappa_{\star\mathrm{L}}(z) + 2.$$

Using the argument in Lemma 8.1, we see that $b_k/c_k = h_k/\tau_k \to 0$. Consequently, for any two points of continuity $x_- < x$ for κ_{L}, as above, we infer that

$$\kappa_{\mathrm{L}}(x_-) \le \liminf \frac{h_k}{\tau_k}\hat{\kappa}_{\mathrm{L}}\left(\frac{x}{h_k}\right) \le \limsup \frac{h_k}{\tau_k}\left(\hat{\kappa}_{\mathrm{L}}\left(\frac{x}{h_k}\right) + 2\right) \le \kappa_{\mathrm{L}}(x).$$

The second limit in (8.5) follows, and this proves the universality of the coarse-grained measure $\mu_\star$. This concludes the existence part of Theorem 8.1.

To prove the density part, we use Lemma 8.1 to note that any probability measure on $\mathbb{N}_0$ with the same tail as $\hat{\pi}$ is also universal. Since all such probability measures are dense in the space of all probability measures on $\mathbb{N}_0$, we obtain the density. This concludes the proof of Theorem 8.1. □

https://doi.org/10.1017/apr.2018.23 Published online by Cambridge University Press

Next, fix a sequence $c_k \to \infty$ that satisfies

$$\sum_{j=1}^{\infty} \frac{j}{c_j} < 1, \qquad c_k \sum_{j>k} \frac{j}{c_j} \to 0 \quad \text{as } k \to \infty.$$

For example, $c_k = c\mathrm{e}^{k^2}$ works for small enough $c > 0$. The main tool used in the proof of Proposition 8.1 is the following 'packing lemma', which is the analog of [33, Lemma 7.2].

Lemma 8.2. (Packing lemma.) *Let $\lambda_k = (0, 0, \mu_k)$ be a sequence of Lévy triples, Φ_k be the corresponding Bernstein transforms*

$$\Phi_k(q) = \int_{\mathbb{R}^+} (1 - \mathrm{e}^{-qx})\, \mathrm{d}\mu_k(x),$$

and assume that

$$\int_{\mathbb{R}^+} \mathrm{d}\mu_k(x) \le k \quad \text{for all } k \in \mathbb{N}. \tag{8.6}$$

Then there exists a sequence $b_k \to \infty$ such that the series

$$\Phi_\star(q) := \sum_{j=1}^{\infty} c_j^{-1} \Phi_j(b_j q) \tag{8.7}$$

converges for each $q > 0$ to a Bernstein function with the following property:

$$c_k \Phi_\star(b_k^{-1} q) - \Phi_k(q) \to 0 \quad \text{as } k \to \infty \text{ for all } q > 0. \tag{8.8}$$

The function $\Phi_\star$ is the Bernstein transform of a Lévy triple of the form $\lambda_\star = (0, 0, \mu_\star)$, where $\mu_\star$ is a finite measure on $(0, \infty)$ with $\int_{\mathbb{R}^+} \mathrm{d}\mu_\star(x) < 1$.

Proof. As $\Phi_j(b_j q) \le \Phi_j(\infty) \le j$ for all $q > 0$, series (8.7) is uniformly bounded and converges for each q. We estimate the quantity in (8.8) in two parts, to show how b_k can be chosen. Note that

$$\left| \Phi_k(q) - c_k \sum_{j=k+1}^{\infty} c_j^{-1} \Phi_j(b_k^{-1} b_j q) \right| \le c_k \sum_{j=k+1}^{\infty} j c_j^{-1} \to 0 \quad \text{as } k \to \infty,$$

regardless of what b_k is. Then, since $\Phi_j(q) \to 0$ as $q \to 0$ for each j, we may choose b_k (inductively) so large that

$$c_k \sum_{j=1}^{k-1} c_j^{-1} \Phi_j(b_k^{-1} b_j q) < \frac{1}{k}.$$

Then (8.8) follows.

The limit function $\Phi_\star$ is the Bernstein transform of some Lévy triple $\lambda_\star = (a_0, a_\infty, \mu_\star)$ by Theorem A.2. Since $\Phi_j(0^+) = 0$, we infer that

$$\Phi_\star(0^+) \le \sum_{j=k+1}^{\infty} j c_j^{-1} \quad \text{for all } k,$$

hence, $a_\infty = \Phi_\star(0^+) = 0$. Furthermore,

$$\Phi_\star(\infty) \le \sum_{j \ge 1} j c_j^{-1} < 1,$$

whence $a_0 = 0$ and $\int_{\mathbb{R}^+} \mathrm{d}\mu_\star(x) < 1$. □

https://doi.org/10.1017/apr.2018.23 Published online by Cambridge University Press

Next we show that Lévy triples of the form $(0, 0, \mu)$ are dense in the space of all Lévy triples. This is analogous to [33, Lemma 7.3].

Lemma 8.3. *Let $\lambda = (a_0, a_\infty, \mu)$ be a Lévy triple. Then there is a sequence of measures μ_k that satisfies (8.6) such that the triples*

$$\lambda_k = (0, 0, \mu_k) \to \lambda \quad \textit{as } k \to \infty.$$

Proof. It suffices to consider λ of the form $\lambda = (0, 0, \mu)$ since these are dense in the space of Lévy triples. But then there exist $\varepsilon_k \to 0$ such that $\mu_k := \mathbf{1}_{\{x \geq \varepsilon_k\}}\mu$ satisfies $\int_{\mathbb{R}_+} \mathrm{d}\mu_k \leq k/3$. Evidently, $\lambda_k = (0, 0, \mu_k) \to \lambda$ due to Definition 8.1. □

With this result we are now ready to prove Proposition 8.1.

Proof of Proposition 8.1. Since the space of finite measures on $[0, \infty]$ with the weak-$\star$ topology is separable, the same also holds for the space of Lévy triples. Thus, we can choose a sequence of Lévy triples $(\bar{\lambda}_n)$ so that *every* Lévy triple λ whatsoever is a limit of (λ_k) along some subsequence.

Partition the integers into infinitely many subsequences and, using Lemma 8.3, select measures μ_k satisfying (8.6) such that $\lambda_k = (0, 0, \mu_k) \to \bar{\lambda}_n$ along the nth subsequence. Construct a sequence b_k, a Bernstein function $\Phi_\star$, and its associated Lévy triple $\lambda_\star$ using Lemma 8.2. We claim that $\lambda_\star$ is universal with sequences (b_k) and (c_k).

To see this, let $\Phi_\star^{b_k,c_k}$ be the Bernstein transform of the rescaled Lévy triple $\lambda_\star^{b_k,c_k}$. Observe that

$$\Phi_\star^{b_k,c_k}(q) = c_k \Phi_\star\left(\frac{q}{b_k}\right). \tag{8.9}$$

Now, given any Lévy triple λ, choose a subsequence (indexed by k) along which $(\bar{\lambda}_k) \to \lambda$. If $\bar{\Phi}_k$ is the Bernstein transform of $\bar{\lambda}_k$ then the continuity theorem for Bernstein transforms (Theorem A.2) guarantees that $(\bar{\Phi}_k) \to \Phi$ pointwise. By (8.8), this implies that $(\Phi_\star^{b_k,c_k}) \to \Phi$ and by the continuity theorem again, we have $(\lambda_\star^{b_k,c_k}) \to \lambda$ along this subsequence. □

9. Linearization and universality for critical CSBPs

The dilational form of the scaling relations (8.2), (8.4), and (8.9) hints at exact scaling relations that hold for the limiting CSBPs, which we develop in this section. Using these relations we establish that the *nonlinear dynamics of CSBPs become linear and purely dilational* when expressed in terms of the Lévy triple that represents the branching mechanism. The map from Lévy triples to finite (sub)critical CSBPs is bicontinuous due to Theorem 9.1 below, and this reduces the study of scaling dynamics for these CSBPs to the study of scaling limits of dilation maps. We use this correspondence here to prove the existence of universal critical CSBPs whose subsequential scaling limits include all possible finite (sub)critical CSBPs.

Our results on linearization here are strongly analogous to the results of Menon and Pego [33], in which the authors showed that scaling dynamics on the scaling attractor for solvable Smoluchowski equations become linear and dilational in terms of a Lévy–Khintchine representation. The definition of the scaling attractor in [33] as the collection of all limits of rescaled solutions was motivated by an analogy to the notion of infinite divisibility in probability. The analogy between infinite divisibility and CSBPs in branching processes has long been evident, at least since the work of Grimvall. But this classical analogy does not appear to explain why the nonlinear dynamics of CSBPs (or the scaling attractor of Smoluchowski equations) become linear in terms of Lévy–Khintchine representations.

9.1. Linearization of renormalized CSBP dynamics

For finite (sub)critical CSBPs, the linearization property we are talking about is stated in the following result. It is actually a simple consequence of the known dynamics of the Laplace exponent in terms of the Lévy–Khintchine representation of the branching mechanism.

Proposition 9.1. *Let $Z(x,t)$ be a CSBP with Laplace exponent $\varphi(q,t)$ and branching mechanism Ψ. Let $b, c > 0$. Then the rescaled CSBP given by*

$$\tilde{Z}(x,t) = b^{-1} Z(bx, ct)$$

satisfies $\tilde{Z}(x,0) = x$ and has Laplace exponent $\tilde{\varphi}$ and branching mechanism $\tilde{\Psi}$ with

$$\tilde{\varphi}(q,t) = b\,\varphi(b^{-1}q, ct), \qquad \tilde{\Psi}(q) = cb\,\Psi(b^{-1}q). \tag{9.1}$$

If the branching mechanism Ψ is (sub)critical, and given by (6.6) in terms of the Lévy triple $(\alpha_0, \alpha_\infty, \mu)$, then the branching mechanism $\tilde{\Psi}$ is determined similarly by the Lévy triple $(\tilde{\alpha}_0, \tilde{\alpha}_\infty, \tilde{\mu})$, where

$$\tilde{\alpha}_0 = cb^{-1}\alpha_0, \qquad \tilde{\alpha}_\infty = c\,\alpha_\infty, \qquad \mathrm{d}\tilde{\mu}(x) = c\,\mathrm{d}\mu(bx), \tag{9.2}$$

corresponding to the left and right distribution functions given by the scaling relations

$$\tilde{\kappa}_{\mathrm{L}}(x) = cb^{-1}\kappa_{\mathrm{L}}(bx), \qquad \tilde{\kappa}_{\mathrm{R}}(x) = c\kappa_{\mathrm{R}}(bx). \tag{9.3}$$

Remark 9.1. Scaling relations (9.2) and (9.3) are exactly the same as scaling relations (8.2) and (8.4) in Section 8.

Proof of proposition 9.1. The first relation in (9.1) follows from the computation

$$\mathrm{e}^{-x\tilde{\varphi}(q,t)} = \mathbb{E}(\mathrm{e}^{-q\tilde{Z}(x,t)}) = \mathbb{E}(\mathrm{e}^{-qb^{-1}Z(bx,ct)}) = \mathrm{e}^{-xb\varphi(b^{-1}q,ct)},$$

and the second relation follows by considering the ordinary differential equation (5.3) satisfied by φ and the corresponding one satisfied by $\tilde{\varphi}$ at $t = 0$, recalling that $\varphi(q,0) = q = \tilde{\varphi}(q,0)$. The relations involving Lévy triples and associated left and right distribution functions follow by comparing the respective representations for Ψ and $\tilde{\Psi}$ from (6.6), and using the definitions in (A.5). □

For the study of long-time scaling limits of CSBPs, we should take b and c as functions of t with $c(t) \to \infty$ or as sequences (b_k) and (c_k) with $c_k \to \infty$. This study can be reduced to the study of scaling limits of the purely dilational relations in (9.3) due to the following continuity theorem. The proof of this theorem is essentially similar to (but simpler than) the proof of Propositions 6.2 and 6.3 for the scaling limits of GW processes.

For brevity in expressing our result, let us say that the Lévy triple $\lambda = (\alpha_0, \alpha_\infty, \mu)$ in (6.6) *generates* the branching mechanism Ψ and the corresponding (finite, subcritical) CSBP Z.

Theorem 9.1. (Continuity theorem for finite (sub)critical CSBPs.) *Let (Z_k) be a sequence of finite (sub)critical CSBPs generated by Lévy triples $\lambda_k = (a_0^k, a_\infty^k, \mu_k)$.*

(i) *Suppose that (λ_k) converges to some Lévy triple λ in the sense of Definition 8.1. Then the finite-dimensional distributions of Z_k converge to those of the finite (sub)critical CSBP Z generated by λ.*

https://doi.org/10.1017/apr.2018.23 Published online by Cambridge University Press

(ii) *Conversely, suppose that the one-dimensional distributions of Z_k converge to those of some finite process Z such that $\mathbb{P}\{Z_{t_0}(x_0) > 0\} > 0$ for some $x_0, t_0 > 0$. Then the sequence λ_k converges to some Lévy triple λ, and Z may be taken as the finite (sub)critical CSBP generated by λ.*

Proof. (i) Suppose that λ_k converges to λ. Then, due to the second continuity theorem (Theorem A.3), the branching mechanisms Ψ_k generated by λ_k converge pointwise to the branching mechanism Ψ generated by λ. Due to representation (A.4) and an argument using Montel's theorem analogously to the proof of Proposition 6.1, the convergence occurs locally uniformly for every derivative. Then it is straightforward to show that since the Laplace exponents φ_k of Z_k satisfy

$$\partial_t \varphi_k(q,t) = -\Psi_k(\varphi_k(q,t)), \qquad \varphi_k(q,0) = q, \tag{9.4}$$

we have $\varphi_k(q,t) \to \varphi(q,t)$ as $k \to \infty$ for each $q > 0$, $t \geq 0$, where φ satisfies (6.16). In particular, φ is the Laplace exponent of the CSBP Z generated by λ. This implies that for each $x, t > 0$, the Laplace transform of $Z_k(x,t)$ converges to that of $Z(x,t)$. This proves the convergence of one-dimensional distributions. The convergence of finite-dimensional distributions follows using arguments similar to Lamperti [29, p. 280].

(ii) Assume that the one-dimensional distributions of Z_k converge to those of some finite process Z such that $\mathbb{P}\{Z(x_0,t_0) > 0\} > 0$ for some $x_0, t_0 > 0$. Then since the Laplace transforms converge, i.e.

$$\mathrm{e}^{-x\varphi_k(q,t)} = \mathbb{E}(\mathrm{e}^{-qZ_k(x,t)}) \to \mathbb{E}(\mathrm{e}^{-qZ(x,t)}) \quad \text{for each } x, t > 0,$$

we have

$$\varphi(q,t) := \lim_{k\to\infty} \varphi_k(q,t)$$

exists for each $q, t > 0$. Taking $t = t_0$, we find that $\varphi(q,t_0) > 0$ for all $q > 0$. For small enough $q > 0$, then, $2q < \varphi(\hat{q}, t_0)$ for some $\hat{q}$, and since Ψ_k is increasing and convex, for large enough k, we have $2q < \varphi_k(\hat{q}, t_0)$, hence, by integrating (9.4), we obtain

$$\Psi_k(2q) \leq \frac{1}{t_0}\int_0^{t_0} \Psi_k(\varphi_k(\hat{q},t))\,\mathrm{d}t \leq \frac{\hat{q}}{t_0}.$$

Hence, with κ_k determined from λ_k as in (8.1) and $g_q(x)$ defined by (A.3), we obtain

$$\langle g_q, \kappa_k\rangle = \Psi_k'(q) \leq q\Psi_k(2q) \leq \frac{q\hat{q}}{t_0}.$$

As in the proof of the continuity theorem (Theorem A.2 (converse part)), this implies that $\langle 1, \kappa_k\rangle$ is bounded. Hence, the sequence of Lévy measures (λ_k) is pre-compact. Along any subsequence that converges to some Lévy triple λ, we may invoke part (i) to assert that the finite-dimensional distributions of Z_k converge to those of the finite (sub)critical CSBP generated by λ. The one-dimensional distributions of this CSBP are then the same as those of Z, independent of the subsequence. It follows that λ is unique, and the whole sequence (λ_k) converges. □

9.2. Existence of universal critical CSBPs

The continuity theorem (Theorem 9.1) could serve as the basis for a comprehensive theory of long-time scaling limits of critical CSBP, but such a study is beyond the scope of this paper. A large number of results exist in the classical literature that cover supercritical and subcritical

https://doi.org/10.1017/apr.2018.23 Published online by Cambridge University Press

cases; see, for example, [2], [18], [25], [27], [40], and [41]. For relatively recent results on critical cases, we refer the reader to [22], [34], and [35]. In [22] we provided necessary and sufficient criteria for the approach to the self-similar form for critical CSBPs that become extinct almost surely (having a branching mechanism satisfying Grey's condition), under a quite general assumption on the scaling taken, but assuming there is a unique limit as $t \to \infty$.

What we will point out here, however, is that the existence of certain *universal* critical CSBPs is now a simple consequence of our study of universal Lévy triples and GW family-size distributions in Section 8. This observation leads to a precise meaning to a remark made by Grey [18] to the effect that a large class of 'critical and subcritical processes . . . do not seem to lend themselves to suitable scaling' which yields a well-defined limit.

Theorem 9.2. *There exists a finite critical CSBP $Z_\star$ that is universal in the following sense. There exist sequences $(b_k), (c_k) \to \infty$ such that, for any finite (sub)critical branching process $\tilde{Z}$, there exists a subsequence along which the finite-dimensional distributions of $Z_\star^{b_k,c_k}$ converge to those of $\tilde{Z}$. Here $Z_\star^{b_k,c_k}$ is the rescaled process defined by*

$$Z_\star^{b_k,c_k}(x,t) := b_k^{-1} Z_\star(b_k x, c_k t).$$

Proof. Using Proposition 8.1, choose $\lambda_\star = (0, 0, \mu_\star)$ to be a universal Lévy triple with sequences $b_k, c_k \to \infty$. Let $Z_\star$ be the critical CSBP generated by $\lambda_\star$. The theorem now follows immediately from the continuity theorem (Theorem 9.1). □

Remark 9.2. Let $\varphi_\star$ and $\tilde{\varphi}$ be the Laplace exponents of $Z_\star$ and $\tilde{Z}$, respectively. Then along the subsequence for which the above convergence holds, we have

$$b_k\, \varphi_\star(b_k^{-1} q, c_k t) \to \tilde{\varphi}(q,t) \quad \text{for all } q, t \in [0,\infty).$$

For each $t > 0$, the functions $\varphi_\star(\cdot, t)$ and $\tilde{\varphi}(\cdot, t)$ are the Bernstein transforms of the respective Lévy triples of the form $(\beta_{\star 0}(t), 0, \nu_{\star,t})$ and $(\tilde{\beta}_0(t), 0, \tilde{\nu}_t)$, with

$$\varphi_\star(q,t) = \beta_{\star 0}(t) q + \int_0^\infty (1 - \mathrm{e}^{-qx})\, \mathrm{d}\nu_{\star,t}(x)$$

and a similar expression for $\tilde{\varphi}$; see (6.17). We note that

$$\beta_{\star 0}(t) = \exp(-\Psi_\star'(\infty) t), \qquad \Psi_\star'(\infty) = \int_0^\infty \mathrm{d}\mu_\star(x) < \infty,$$

and $\beta_{\star 0}(c_k t) \to 0$ as $k \to \infty$. Due to the continuity theorem (Theorem A.2), the convergence above corresponds to a Lévy-convergence property: for each $t > 0$, as $k \to \infty$,

$$(x \wedge 1) b_k\, \mathrm{d}\nu_{\star, c_k t}(b_k x) \to \beta_0(t)\delta_0 + (x \wedge 1)\, \mathrm{d}\tilde{\nu}_t(x) \quad \text{weak-}\star \text{ on } [0,\infty].$$

Appendix A. Continuity theorems for Bernstein transforms of Lévy triples

The Lévy-convergence requirement (Definition 6.1) used in all our results relates to a natural topology of Lévy triples associated with subordinators. Our purpose in this appendix is to expand on this and prove a couple of continuity theorems relating convergence of Lévy triples to pointwise convergence of the associated Bernstein functions. These theorems are variants of the classical continuity theorem for Laplace transforms, but do not appear to be widely known.

https://doi.org/10.1017/apr.2018.23 Published online by Cambridge University Press

Definition A.1. We say that (a_0, a_∞, μ) is a *Lévy triple* if μ is a (nonnegative) measure on $\mathbb{R}^+ = (0, \infty)$, and

$$a_0 \geq 0, \qquad a_\infty \geq 0, \qquad \int_{\mathbb{R}^+} (x \wedge 1) \, \mathrm{d}\mu(x) < \infty.$$

A measure μ satisfying the above is called a *Lévy measure*.

For the next definition, recall that a smooth function $g \colon \mathbb{R}^+ \to \mathbb{R}$ is said to be *completely monotone* if $(-1)^n g^{(n)} \geq 0$ for all integer $n \geq 0$.

Definition A.2. A function $f \colon \mathbb{R}^+ \to \mathbb{R}$ is *Bernstein* if it is smooth, nonnegative, and its derivative f' is completely monotone.

Schilling *et al.* [38] developed the theory of Bernstein functions extensively. The main representation theorem regarding these functions (see Theorem 3.2 of [38]) is the following variant of Bernstein's theorem (which states that g is completely monotone if and only if it is the Laplace transform of some Radon measure on $[0, \infty)$).

Theorem A.1. *A function $f \colon \mathbb{R}^+ \to \mathbb{R}$ is Bernstein if and only if it has the representation*

$$f(q) = a_0 q + a_\infty + \int_{\mathbb{R}^+} (1 - \mathrm{e}^{-qx}) \, \mathrm{d}\mu(x) \quad \textit{for some Lévy triple } (a_0, a_\infty, \mu). \tag{A.1}$$

In particular, the triple (a_0, a_∞, μ) determines f uniquely and vice versa.

For convenience, we call the function f in (A.1) the *Bernstein transform* of the Lévy triple (a_0, a_∞, μ). If $a_0 = a_\infty = 0$, we call f the Bernstein transform of μ.

Lévy triples of the form above arise naturally in the study of subordinators—right-continuous, increasing (possibly infinite) processes that have independent, time-homogeneous increments. We know (see, for example, [8]) that the Laplace exponent of a subordinator can be uniquely expressed in the form (A.1), as the Bernstein transform of some Lévy triple. In terms of the subordinator, a_0 represents the drift, a_∞ the killing, and μ the jumps.

We obtain a natural topology on Lévy triples by associating to each Lévy triple a finite measure on the compactified half-line $[0, \infty]$, including atoms at 0 and ∞. Explicitly, to any Lévy triple (a_0, a_∞, μ), we associate the finite measure κ on $[0, \infty]$ defined by

$$\mathrm{d}\kappa(x) = a_0 \, \mathrm{d}\delta_0(x) + a_\infty \, \mathrm{d}\delta_\infty(x) + (x \wedge 1) \, \mathrm{d}\mu(x). \tag{A.2}$$

We note that this association of finite measures with Lévy triples is bijective. Now we use the weak-$\star$ topology of finite measures on $[0, \infty]$ to induce a topology on the set of Lévy triples. Note that, for any $g \in C([0, \infty])$, we have

$$\langle g, \kappa \rangle = g(0) a_0 + g(\infty) a_\infty + \int_{\mathbb{R}^+} g(x)(x \wedge 1) \, \mathrm{d}\mu(x).$$

Definition A.3. We say that a sequence of Lévy triples $(a_0^{(k)}, a_\infty^{(k)}, \mu_k)$ converges to the Lévy triple (a_0, a_∞, μ) if the corresponding sequence of measures κ_k converges to κ weak-$\star$ on $[0, \infty]$. That is, for every $g \in C([0, \infty])$, we have

$$\langle g, \kappa_k \rangle \to \langle g, \kappa \rangle \quad \text{as } k \to \infty.$$

https://doi.org/10.1017/apr.2018.23 Published online by Cambridge University Press

Remark A.1. This is exactly the same as Definition 8.1, restated here for convenience. Moreover, this generalizes the notion of Lévy-convergence introduced in Definition 6.1. Indeed, Lévy-convergence of $(\breve{\mu}_k)$ to a measure κ is exactly the convergence of the Lévy triples $(0, 0, \breve{\mu}_k)$ to the associated Lévy triple $(\kappa(0), \kappa(\infty), \kappa|_{(0,\infty)})$.

In what follows we provide a continuity theorem for the Bernstein transform that is not present in [38]. It may be inferred from the proof of [33, Theorem 3.1], but [33] employs a different and rather cumbersome terminology, and the proof below is much simpler.

Theorem A.2. *Let $(a_0^{(k)}, a_\infty^{(k)}, \mu_k)$ be a sequence of Lévy triples with Bernstein transforms f_k corresponding via (A.1) and measures κ_k corresponding via (A.2). Then the following are equivalent:*

(i) $f(q) := \lim_{k\to\infty} f_k(q)$ *exists for each* $q \in (0, \infty)$*;*

(ii) κ_k *converges to some finite measure* κ *weak-*$\star$ *on* $[0, \infty]$*. That is, as* $k \to \infty$*,*

$$\langle g, \kappa_k\rangle \to \langle g, \kappa\rangle \quad \textit{for every } g \in C([0, \infty]).$$

If either condition holds then the limits f and κ correspond to a unique Lévy triple via (A.1) and (A.2), respectively.

By this result, the sequence (f_k) converges pointwise on $\mathbb{R}^+$ if and only if the sequence of Lévy triples $(a_0^{(k)}, a_\infty^{(k)}, \mu_k)$ converges in the sense of Definition A.3.

Proof of Theorem A.2. Let (κ_k) be the sequence of measures associated with the Lévy triples $(a_0^{(k)}, a_\infty^{(k)}, \mu_k)$ as in (A.2). Suppose first that $\kappa_k \to \kappa$ weak-$\star$ on $[0, \infty]$. For $q \in \mathbb{R}^+$, consider the test function $g_q : \mathbb{R}^+ \to \mathbb{R}$ defined by

$$g_q(x) = \frac{1 - \mathrm{e}^{-qx}}{x \wedge 1}. \tag{A.3}$$

By defining $g_q(0) = q$ and $g_q(\infty) = 1$, we can extend g_q to a continuous function on $[0, \infty]$. Consequently,

$$\langle g_q, \kappa\rangle = \lim_{k\to\infty} \langle g_q, \kappa_k\rangle = \lim_{k\to\infty} a_0^{(k)} q + a_\infty^{(k)} + \int_{\mathbb{R}^+} (1 - \mathrm{e}^{-qx})\, \mathrm{d}\mu(x) = \lim_{k\to\infty} f_k(q),$$

establishing pointwise convergence of (f_k) on $\mathbb{R}^+$, as desired.

For the converse, suppose that $(f_k) \to f$ pointwise on $\mathbb{R}^+$. Note that, for $q \in \mathbb{R}^+$, we have

$$c_q := \inf_{x\in\mathbb{R}^+} g_q(x) > 0.$$

Consequently,

$$\sup_k \langle 1, \kappa_k\rangle \le \sup_k \frac{1}{c_q} \langle g_q, \kappa_k\rangle = \frac{1}{c_q} \sup_k f_k(q) < \infty.$$

Thus, by the Banach–Alaoglu theorem, any subsequence of (κ_k) has a further subsequence that is weak-$\star$ convergent on $[0, \infty]$. Let κ denote any such subsequential limit. By taking limits as above but along subsequences, we infer that, for every $q \in \mathbb{R}^+$,

$$\langle g_q, \kappa\rangle = f(q).$$

This shows that f is the Bernstein transform of the Lévy triple associated to κ by (A.2). Since both this association and the Bernstein transform are bijective, κ is uniquely determined by f. It follows that the entire sequence (κ_k) converges weak-$\star$ to the same limit κ. □

https://doi.org/10.1017/apr.2018.23 Published online by Cambridge University Press

We finish by developing two further variations of the convergence conditions in the continuity theorem above. To each Lévy triple (a_0, a_∞, μ), we associate two further quantities:

- the *Bernstein primitive* (or branching mechanism)

$$\psi(q) = \frac{1}{2}a_0 q^2 + a_\infty q + \int_{\mathbb{R}^+} \frac{\mathrm{e}^{-qx} - 1 + qx}{x}\,\mathrm{d}\mu(x); \tag{A.4}$$

- the pair of *left and right distribution functions* $(\kappa_\mathrm{L}, \kappa_\mathrm{R})$ given by

$$\kappa_\mathrm{L}(x) = a_0 + \int_{(0,x]} z\,\mathrm{d}\mu(z), \qquad \kappa_\mathrm{R}(x) = a_\infty + \int_{(x,\infty)} \mathrm{d}\mu(z). \tag{A.5}$$

These distribution functions are associated to Radon measures, also denoted by κ_L on $[0, \infty)$ and κ_R on $(0, \infty]$, in a standard way.

Theorem A.3. *We make the same assumptions as in Theorem A.2 and, for each k, associate* ψ_k *and* $(\kappa_{\mathrm{L},k}, \kappa_{\mathrm{R},k})$ *via (A.4) and (A.5), respectively. Then the following conditions are also equivalent to (i) and (ii) of Theorem A.2.*

(iii) $\psi(q) := \lim_{k\to\infty} \psi_k(q)$ *exists for each* $q \in (0, \infty)$.

(iv) *For almost every* $x \in (0, \infty)$, *both of the following limits exist:*

$$\kappa_\mathrm{L}(x) = \lim_{k\to\infty} \kappa_{\mathrm{L},k}(x), \qquad \kappa_\mathrm{R}(x) = \lim_{k\to\infty} \kappa_{\mathrm{R},k}(x).$$

Remark A.2. Condition (iv) is equivalent to weak-$\star$ convergence of the measures $\kappa_{\mathrm{L},k}$ and $\kappa_{\mathrm{R},k}$ on the intervals $[0, \infty)$ and $(0, \infty]$, respectively.

Proof of Theorem A.3. It suffices to show that Theorem A.2(i) is equivalent to (iii), and Theorem A.2(ii) is equivalent to (iv).

Suppose that Theorem A.2(i) holds. Note that f_k is increasing and concave for all k, hence, for each $q \in \mathbb{R}^+$, by dominated convergence, we have

$$\psi_k(q) = \int_0^q f_k(r)\,\mathrm{d}r \to \int_0^q f(r)\,\mathrm{d}r =: \psi(q).$$

Hence, (iii) holds.

Conversely, suppose that (iii) holds. Note that $\psi_k(2q) \ge q f_k(q)$ for all $q > 0$. For each $q > 0$, it follows that $(f_k(q))$ is pre-compact. Since f_k is increasing and concave, each subsequence has a further subsequence that converges pointwise to some (increasing and concave) function f. Necessarily,

$$\int_0^q f(r)\,\mathrm{d}r = \psi(q) \quad \text{for each } q > 0,$$

therefore, ψ is C^1 and $\psi' = f$. Thus, f is determined by ψ, and it follows that the whole sequence (f_k) converges, proving (i).

Next we prove that Theorem A.2(ii) implies (iv). Given κ as in Theorem A.2(ii), we may define $\kappa_{\rm L}, \kappa_{\rm R}$ so that

$$\mathrm{d}\kappa_{\rm L}(x) = \frac{x\,\mathrm{d}\kappa(x)}{x \wedge 1} \quad \text{and} \quad \mathrm{d}\kappa_{\rm R}(x) = \frac{\mathrm{d}\kappa(x)}{x \wedge 1}$$

on the intervals $[0, \infty)$ and $(0, \infty]$, respectively. For any $g_0 \in C_c([0, \infty))$ and $g_\infty \in C_c((0, \infty])$, let

$$g_{\rm L}(x) = \frac{x\, g_0(x)}{x \wedge 1} \quad \text{and} \quad g_{\rm R}(x) = \frac{g_\infty(x)}{x \wedge 1}. \tag{A.6}$$

Then, as $k \to \infty$, we have

$$\langle g_0, \kappa_{{\rm L},k}\rangle = \langle g_{\rm L}, \kappa_k\rangle \to \langle g_{\rm L}, \kappa\rangle =: \langle g_0, \kappa_{\rm L}\rangle,$$
$$\langle g_\infty, \kappa_{{\rm R},k}\rangle = \langle g_{\rm R}, \kappa_k\rangle \to \langle g_{\rm R}, \kappa\rangle =: \langle g_\infty, \kappa_{\rm R}\rangle.$$

This establishes weak-$\star$ convergence of the measures $\kappa_{{\rm L},k}$ to $\kappa_{\rm L}$ and $\kappa_{{\rm R},k}$ to $\kappa_{\rm R}$; hence, (iv) holds.

Conversely, assume that (iv) holds. Fix smooth cutoff functions $\upsilon_{\rm L}, \upsilon_{\rm R} \colon [0, \infty] \to [0, 1]$ such that $\upsilon_{\rm L} + \upsilon_{\rm R} = 1$ and $\upsilon_{\rm L} = 1$ on $[0, 1]$, and $\upsilon_{\rm R} = 1$ on $[2, \infty]$. Given any $g \in C([0, \infty])$, we write $g = g_{\rm L} + g_{\rm R}$, where $g_{\rm L} = g\upsilon_{\rm L}$ and $g_{\rm R} = g\upsilon_{\rm R}$, and determine g_0 and g_∞ by (A.6). Then

$$\langle g_{\rm L} + g_{\rm R}, \kappa_k\rangle = \langle g_0, \kappa_{{\rm L},k}\rangle + \langle g_\infty, \kappa_{{\rm R},k}\rangle \to \langle g_0, \kappa_{\rm L}\rangle + \langle g_\infty, \kappa_{\rm R}\rangle \quad \text{as } k \to \infty. \tag{A.7}$$

Since $\lim_{k\to\infty}\langle g, \kappa_k\rangle$ exists for each g, there exists a measure κ on $[0, \infty]$ with $\kappa_k \to \kappa$ weak-$\star$ on $[0, \infty]$. Thus, Theorem A.2(ii) holds. □

Acknowledgements

This material is based upon work partially supported by the National Science Foundation (grant numbers DMS-1252912 and DMS-1515400), the Simons Foundation (grant numbers 393685 and 395796), and the Center for Nonlinear Analysis (grant number NSF OISE-0967140).

References

[1] Aldous, D. J. (1999). Deterministic and stochastic models for coalescence (aggregation and coagulation): a review of the mean-field theory for probabilists. *Bernoulli* **5,** 3–48.
[2] Athreya, K. B. and Ney, P. E. (2004). *Branching Processes*. Dover, Mineola, NY.
[3] Bacaër, N. (2011). *A Short History of Mathematical Population Dynamics*. Springer, London.
[4] Bansaye, V. and Simatos, F. (2015). On the scaling limits of Galton-Watson processes in varying environments. *Electron. J. Prob.* **20,** 75.
[5] Berestycki, N. (2009). *Recent Progress in Coalescent Theory* (Math. Surveys **16**). Sociedade Brasileira de Matemática, Rio de Janeiro.
[6] Berestycki, J., Berestycki, N., and Limic, V. (2014). A small-time coupling between Λ-coalescents and branching processes. *Ann. Appl. Prob.* **24,** 449–475.
[7] Berestycki, J., Berestycki, N. and Schweinsberg, J. (2008). Small-time behavior of beta coalescents. *Ann. Inst. H. Poincaré Prob. Statist.* **44,** 214–238.
[8] Bertoin, J. (2000). Subordinators, Lévy processes with no negative jumps, and branching processes. Preprint. Université Pierre et Marie Curie.
[9] Bertoin, J. (2006). *Random Fragmentation and Coagulation Processes* (Camb. Studies Adv. Math. **102**). Cambridge University Press.
[10] Bertoin, J. and Le Gall, J.-F. (2000). The Bolthausen-Sznitman coalescent and the genealogy of continuous-state branching processes. *Prob. Theory Relat. Fields* **117,** 249–266.

[11] Bertoin, J. and Le Gall, J.-F. (2003). Stochastic flows associated to coalescent processes. *Prob. Theory Relat. Fields* **126,** 261–288.
[12] Bertoin, J. and Le Gall, J.-F. (2005). Stochastic flows associated to coalescent processes. II. Stochastic differential equations. *Ann. Inst. H. Poincaré Prob. Statist.* **41,** 307–333.
[13] Bertoin, J. and Le Gall, J.-F. (2006). Stochastic flows associated to coalescent processes. III. Limit theorems. *Illinois J. Math.* **50,** 147–181.
[14] Birkner, M. *et al.* (2005). Alpha-stable branching and beta-coalescents. *Electron. J. Prob.* **10,** 303–325.
[15] Caballero, M. E., Lambert, A. and Uribe Bravo, G. (2009). Proof(s) of the Lamperti representation of continuous-state branching processes. *Prob. Surveys* **6,** 62–89.
[16] Feller, W. (1971). *An Introduction to Probability Theory and Its Applications,* Vol. II, 2nd edn. John Wiley, New York.
[17] Gnedin, A., Iksanov, A. and Marynych, A. (2014). Λ-coalescents: a survey. In *Celebrating 50 Years of The Applied Probability Trust* (J. Appl. Prob. **51A**), pp. 23–40.
[18] Grey, D. R. (1974). Asymptotic behaviour of continuous time, continuous state-space branching processes. *J. Appl. Prob.* **11,** 669–677.
[19] Grimvall, A. (1974). On the convergence of sequences of branching processes. *Ann. Prob.* **2,** 1027–1045.
[20] Grosjean, N. and Huillet, T. (2016). On a coalescence process and its branching genealogy. *J. Appl. Prob.* **53,** 1156–1165.
[21] Heyde, C. C. and Seneta, E. (1977). *I. J. Bienaymé: Statistical Theory Anticipated*. Springer, New York.
[22] Iyer, G., Leger, N. and Pego, R. L. (2015). Limit theorems for Smoluchowski dynamics associated with critical continuous-state branching processes. *Ann. Appl. Prob.* **25,** 675–713.
[23] Kingman, J. F. C. (1982). The coalescent. *Stoch. Process. Appl.* **13,** 235–248.
[24] Kingman, J. F. C. (1982). On the genealogy of large populations. In *Essays in Statistical Science* (J. Appl. Prob. **19A**), pp. 27–43.
[25] Kyprianou, A. E. (2014). *Fluctuations of Lévy Processes with Applications*, 2nd edn. Springer, Heidelberg.
[26] Lambert, A. (2003). Coalescence times for the branching process. *Adv. Appl. Prob.* **35,** 1071–1089.
[27] Lambert, A. (2007). Quasi-stationary distributions and the continuous-state branching process conditioned to be never extinct. *Electron. J. Prob.* **12,** 420–446.
[28] Lamperti, J. (1967). Continuous state branching processes. *Bull. Amer. Math. Soc.* **73,** 382–386.
[29] Lamperti, J. (1967). The limit of a sequence of branching processes. *Z. Wahrscheinlichkeitsth.* **7,** 271–288.
[30] Laurençot, P. and van Roessel, H. (2015). Absence of gelation and self-similar behavior for a coagulation-fragmentation equation. *SIAM J. Math. Anal.* **47,** 2355–2374.
[31] Li, Z.-H. (2000). Asymptotic behaviour of continuous time and state branching processes. *J. Austral. Math. Soc. Ser. A* **68,** 68–84.
[32] Menon, G. and Pego, R. L. (2004). Approach to self-similarity in Smoluchowski's coagulation equations. *Commun. Pure Appl. Math.* **57,** 1197–1232.
[33] Menon, G. and Pego, R. L. (2008). The scaling attractor and ultimate dynamics for Smoluchowski's coagulation equations. *J. Nonlinear Sci.* **18,** 143–190.
[34] Pakes, A. G. (2008). Conditional limit theorems for continuous time and state branching process. In *Records and Branching Processes*, Nova, New York, pp. 63–103.
[35] Pakes, A. G. (2010). Critical Markov branching process limit theorems allowing infinite variance. *Adv. Appl. Prob.* **42,** 460–488.
[36] Pitman, J. (1999). Coalescents with multiple collisions. *Ann. Prob.* **27,** 1870–1902.
[37] Sagitov, S. (1999). The general coalescent with asynchronous mergers of ancestral lines. *J. Appl. Prob.* **36,** 1116–1125.
[38] Schilling, R. L., Song, R. and Vondraček, Z. (2010). *Bernstein Functions* (De Gruyter Studies Math. **37**). De Gruyter, Berlin.
[39] Schweinsberg, J. (2003). Coalescent processes obtained from supercritical Galton-Watson processes. *Stoch. Process. Appl.* **106**, 107–139.
[40] Vatutin, V. A. and Zubkov, A. M. (1985). Branching processes. I. In *Probability Theory: Mathematical Statistics: Theoretical Cybernetics*, Akad. Nauk SSSR, Moscow, pp. 3–67.
[41] Vatutin, V. A. and Zubkov, A. M. (1993). Branching processes. II. *J. Soviet Math.* **67,** 3407–3485.
[42] Von Smoluchowski, M. (1916). Drei Vorträge über Diffusion, Brownsche Bewegung und Koagulation von Kolloidteilchen. *Physik. Z.* **17,** 557–585.
[43] Von Smoluchowski, M. (1918). Experiments on a mathematical theory of kinetic coagulation of coloid solutions. *Z. Physikalische Chem. Stoch. Verwandtschaftslehre* **92,** 129–168.
[44] Watson, H. W. and Galton, F. (1875). On the probability of the extinction of families. *J. Anthropological Inst. Great Britain Ireland* **4,** 138–144.