

COAGULATION AND UNIVERSAL SCALING LIMITS FOR CRITICAL GALTON-WATSON PROCESSES

GAUTAM IYER¹, NICHOLAS LEGER², AND ROBERT L. PEGO¹

Dedicated to the memory of Jack Carr.

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 necessary and sufficient criteria for the convergence of scaling limits of Galton-Watson processes that are simpler (and equivalent) to the classical criteria obtained by Grimvall in 1974. Our results provide a clear and natural interpretation, and an alternate proof, of the fact that the Lévy jump measure of certain CSBPs satisfies a generalized Smoluchowski equation. (This result was previously proved by Bertoin and Le Gall in 2006.)

Moreover, our analysis shows that the nonlinear scaling dynamics of CSBPs becomes *linear and purely dilatational* when expressed in terms of the Lévy triple associated with the branching mechanism. We use this to prove existence of *universal* critical Galton-Watson and CSBPs analogous to W. 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). We develop these in a self-contained appendix.

¹ DEPARTMENT OF MATHEMATICAL SCIENCES, CARNEGIE MELLON UNIVERSITY, PITTSBURGH, PA 15213.

² DEPARTMENT OF MATHEMATICS, UNIVERSITY OF HOUSTON, 4800 CALHOUN RD., HOUSTON, TX 77004.

E-mail addresses: gautam@math.cmu.edu, nleger@math.uh.edu, rpego@cmu.edu.

2010 *Mathematics Subject Classification.* Primary 60J80; Secondary 37E20, 35Q70.

Key words and phrases. Smoluchowski equation, multiple coalescence, Galton-Watson processes, CSBPs, universal laws, Bernstein transform.

This material is based upon work partially supported by the National Science Foundation under grants the National Science Foundation (through grants DMS-1252912 to GI, DMS-1515400 to RP), the Simons Foundation (through grants #393685 to GI, #395796 to RP), and the Center for Nonlinear Analysis (through grant NSF OISE-0967140).

CONTENTS

1. Introduction.	3
Summary of results.	4
Part 1. Time and size discrete coalescence and branching.	5
2. Coagulation equations and processes with multiple mergers.	5
2.1. A discrete-time process modeling coalescence.	6
2.2. Evolution of the cluster-size distribution.	7
2.3. The discrete Smoluchowski equation.	8
3. Galton-Watson processes and coagulation dynamics.	9
3.1. A brief introduction to Galton-Watson processes	9
3.2. Discrete coagulation dynamics of the descendant distribution.	10
4. Bernstein transform of the discrete evolution equations	11
Part 2. Convergence criteria governing scaling limits of Galton-Watson processes.	12
5. CSBPs and the time continuous Smoluchowski equation.	12
6. Scaling limits of critical Galton-Watson processes.	15
6.1. Rescaled time-discrete dynamics.	16
6.2. Convergence of critical branching mechanisms.	17
6.3. Convergence of Bernstein transforms	20
6.4. Convergence of Galton-Watson processes.	23
7. Grimvall's convergence criterion, simplified and compared.	25
7.1. A simplification of Grimvall's result	25
7.2. Equivalence in the critical case	27
Part 3. Universality in Galton-Watson process and CSBPs.	30
8. Universal Galton-Watson family size distributions	30
9. Linearization and universality for critical CSBP	35
9.1. Linearization of renormalized CSBP dynamics	35
9.2. Existence of universal critical CSBPs	37
Appendix A. Continuity theorems for Bernstein transforms of Lévy triples.	38
References	42

1. INTRODUCTION.

In 1875, Galton and Watson [42] took up 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 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 processes or Bienaymé-Galton-Watson processes, since key aspects of the topic were discovered thirty years earlier by I. J. Bienaymé (see [3, 19]). 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 [21, 22] 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 sizes of groups or clusters of individuals with a common living ancestor as time evolves backwards. Kingman was particularly concerned with models that emerged from early work of Moran and Wright that hold total population fixed. His work was subsequently expanded to study a general class of exchangeable coalescent models, and a natural completion—the class of Λ -coalescents—was identified and studied by Pitman [34] and Sagitov [35]. We refer to the book [8] and the survey papers [6, 16] for an overview of the now-extensive literature in this area. For aspects particularly related to CSBPs, we refer to [4, 5, 9, 13, 24, 37].

The present work focuses on classic Galton-Watson processes and their continuum limits—continuous state branching processes (CSBPs)—which are models in which populations may fluctuate. Our work was originally motivated by observations of Bertoin and Le Gall contained in a series of papers that relate certain coalescence models to CSBPs [10–12]. In [12] the authors prove a striking result which shows that the Lévy jump measure associated to certain CSBPs satisfies a generalized Smoluchowski coagulation equation with multiple coalescence. One goal of our work is to provide a simple and natural explanation for this fact.

Our starting point for such an explanation is an elementary observation showing that the n -step descendant distribution of the Galton-Watson process satisfies a discrete Smoluchowski coagulation equation with multiple coalescence. We use this basic connection to study scaling limits of Galton-Watson process and the coagulation dynamics of the Lévy jump measure of certain CSBPs.

From our scaling and limit analysis of the coagulation dynamics associated with Galton-Watson processes follows another simple yet striking observation: The *nonlinear scaling dynamics of CSBPs becomes linear and purely dilational* when expressed in terms of the Lévy triple that represents the branching mechanism! This

observation is parallel to results of [31] 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 [18] classifying all continuum limits of Galton-Watson processes as CSBPs. Here we make use of the dilational representation to establish the existence of certain *universal* critical Galton-Watson and CSBPs which generate *all* critical and subcritical CSBPs as subsequential scaling limits. These universal processes are analogous to W. Doeblin’s “universal laws” in classical probability theory [15, XVII.9], and add a precise explanation to a remark made by Grey [17] 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.

Summary of results. The two principal new results we prove in this paper are the existence of universal Galton-Watson processes, and the existence of universal CSBPs.

Theorem 1.1. *There exists a critical Galton-Watson 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. Our paper also contains a variety of other results concerning these issues. We devote the rest of this section to summarizing the main results and organization of our paper.

Time and size discrete coalescence and branching. Part 1 (Sections 2–4) is devoted to the study of discrete-time, discrete-size coalescence and branching. We begin by introducing (Section 2) a time-discrete Markov process C modeling sizes of clusters undergoing coalescence. We show (Proposition 2.3) that the backward equation of the process C is exactly a time-discrete analog of the well known Smoluchowski coagulation equation [40, 41], a mean-field rate-equation model of clustering. Moreover we show (Proposition 3.1) that the one dimensional distributions of coordinates of C have the same distribution as a Galton-Watson process. This provides a elementary and fundamental connection between branching and coalescence, and shows that the n^{th} generation descendant distribution of a critical Galton-Watson process is itself a solution of the discrete Smoluchowski equation with multiple coalescence. This will be used extensively in the scaling analysis performed in Part 2, below.

Finally, we conclude our treatment of the discrete scenario by introducing the *branching mechanism* and the *Bernstein transform* for Galton-Watson processes. We show (Proposition 4.2) that the discrete coagulation dynamics gives 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 [12, 20] or equation (5.5), below), and is a direct analog of the backward equation for critical CSBPs that become extinct almost surely.

Convergence criteria governing scaling limits of Galton-Watson processes. Part 2 (Sections 5–7) is devoted to studying scaling limits of Galton-Watson processes. We begin with a brief introduction to CSBPs (Section 5) and discuss Bertoin and Le Gall’s striking result from [12] showing that the Lévy jump measure of certain CSBPs satisfies a special form of the Smoluchowski equation with multiple coalescence. One of our goals is to explain how this naturally comes to be.

Using improvements in the theory of Bernstein functions [36] (known in probability as Laplace exponents), we obtain simple and precise criteria for the existence of scaling limits of Galton-Watson processes expressed directly in terms of rescaled reproduction laws (Propositions 6.3, 6.4 and 6.6). 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 Galton-Watson processes, and explains how the generalized Smoluchowski equation arises naturally in [12]. 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 separately¹ in Appendix A.

Finally, the convergence of scaling limits of Galton-Watson processes is a classical subject and has been studied by many authors. For completeness, we show (Proposition 7.4) that the convergence criteria we develop in this paper are equivalent to the classical ones provided by [18].

Universality in Galton-Watson processes and CSBPs. Part 3 (Sections 8–9) is devoted to studying universality and proving Theorems 1.1 and 1.2. For Galton-Watson processes, we use our simplified convergence results to show that there exists a Galton-Watson process with a “universal” family size distribution (Theorem 8.1). Namely, we show that there exists a Galton-Watson process X such that *any* (sub)critical CSBP that remains finite almost surely can be obtained by appropriately rescaling X and taking a subsequential limit.

Moreover, our analysis of the coagulation dynamics associated with Galton-Watson processes yields a simple and striking observation regarding the scaling dynamics of CSBPs. Namely, the scaling dynamics of CSBPs becomes *linear and purely dilatational* when expressed in terms of the Lévy triple associated with the branching mechanism (Section 9). We show that the map from Lévy triple to solution is bicontinuous in a particular sense (Proposition 9.1), and 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 9.4).

Part 1. Time and size discrete coalescence and branching.

2. COAGULATION EQUATIONS AND PROCESSES WITH MULTIPLE MERGERS.

The Smoluchowski coagulation equation [40, 41] 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

¹The 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.6).

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 Galton-Watson processes that will be delineated in the following section.

2.1. A discrete-time process modeling coalescence. Let $\hat{\pi}$ be a probability measure on $\mathbb{N}_0 \stackrel{\text{def}}{=} \mathbb{N} \cup \{0\}$. 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.

- (1) 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.
- (2) The probability that a merger involves exactly k clusters is $\hat{\pi}(k)$. We refer to $\hat{\pi}$ as the *merger distribution*.
- (3) The sizes of each of the k clusters participating in a simultaneous k -merger are independent and identically distributed.

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 $C = (C_n)_{n \in \mathbb{N}_0}$ that has the above features proceeds as follows. For each time $n \in \mathbb{N}_0$, C_n is to be 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 independent and identically distributed; however, for the sake of comparison with Galton-Watson processes, here we will always assume

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

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_n for $0 \leq m \leq n$, such that $M_{0,n} = 0$ and the increments

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

are independent and identically distributed with law $\hat{\pi}$. Define

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

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 zero). 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 $n \geq 0$, the random variables $C_n(i)$, $i \in \mathbb{N}$, are independent and identically distributed. Thus (2.1) guarantees that for any $j \in \mathbb{N}$, the random variable $C_{n+1}(j)$ is the sum of N independent and identically distributed 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 (1)–(3) 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. In case $\hat{\pi}(0) = 0$, 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 $n \geq 0$. Our coagulation process corresponds to a discrete-time type of *thinning* of these processes—marks disappear according to the rules that govern the process C . It was pointed out by Aldous in [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 Galton-Watson processes to appear in Section 3 below.

2.2. Evolution of the cluster-size distribution. We 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. Because the variables $\{C_n(i)\}_{i \in \mathbb{N}}$ are all identically distributed, we have

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

Our first observation is that the n -step size distribution ν_n determines ν_{n+1} in a manner that naturally captures (1)–(3).

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

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

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

$$(2.4) \quad \nu_n^{*k}(j) = \sum_{i_1, \dots, i_k \geq 0} \delta^j(i_1 + \dots + i_k) \nu_n(i_1) \cdots \nu_n(i_k),$$

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 formula 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 $\hat{\pi}$. \square

Remark. 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).

Remark. If the expected number of clusters in a simultaneous merger (given by $\sum k \hat{\pi}(k)$) and the expected initial cluster size are both finite, then

$$(2.5) \quad \nu_n(j) = \lim_{N \rightarrow \infty} \frac{|\{k \mid C_n(k) = j \ \& \ 1 \leq k \leq N\}|}{N} \quad \text{almost surely.}$$

Indeed, the above assumptions and induction immediately guarantee $\mathbf{E}C_n(i) < \infty$ for all $i \in \mathbb{N}$. Since the random variables $\{C_n(i)\}_{i \in \mathbb{N}}$ are independent and identically distributed (and have finite first moment) the law of large numbers implies (2.5).

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

Definition 2.2. We say the merger distribution $\hat{\pi}$ is *critical* if

$$(2.6) \quad \sum_{k \geq 0} k \hat{\pi}(k) = 1.$$

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

Proposition 2.3. *If the merger distribution $\hat{\pi}$ is critical, then the size distribution ν_n satisfies*

$$(2.7) \quad \nu_{n+1}(j) - \nu_n(j) = \sum_{k \geq 2} \hat{R}_k(\rho_n) \hat{I}_k(\nu_n, j), \quad j > 0,$$

where

$$(2.8) \quad \rho_n \stackrel{\text{def}}{=} \mathbf{P}\{C_n(1) > 0\} = \sum_{j > 0} \nu_n(j),$$

$$(2.9) \quad \hat{R}_k(\rho_n) \stackrel{\text{def}}{=} \sum_{l \geq k} \hat{\pi}(l) \binom{l}{k} \rho_n^k (1 - \rho_n)^{l-k},$$

$$(2.10) \quad \hat{I}_k(\nu_n, j) \stackrel{\text{def}}{=} \sum_{i_1, \dots, i_k > 0} \left[\delta^j \left(\sum_{l=1}^k i_l \right) - \sum_{l=1}^k \delta^j(i_l) \right] \frac{\nu_n(i_1)}{\rho_n} \dots \frac{\nu_n(i_k)}{\rho_n}.$$

Remark. Equations (2.7)–(2.10) are a discrete version of the Smoluchowski equations (see Section 5, below) 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 non-zero size is given by $\hat{R}_k(\rho_n)$. Now in the event of a simultaneous merger of k non-zero sized clusters, the chance that a cluster of size j is created is exactly the positive term in (2.10). 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 (2.10). 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 non-zero sized clusters occurred. Assuming independence and summing yields (2.7).

Proof of Proposition 2.3. Equation (2.7) follows when we restrict the size dynamics to clusters of non-zero size. Indeed, $\nu_n(j)/\rho_n$ is exactly the chance that a cluster has size j , conditioned on having non-zero size. Since \hat{R}_k is the chance that exactly k such clusters merge, we must have

$$(2.11) \quad \begin{aligned} \nu_{n+1}(j) &= \sum_{k \geq 1} \hat{R}_k(\rho_n) \left(\frac{\nu_n}{\rho_n} \right)^{*k}(j) \\ &= \sum_{k \geq 1} \hat{R}_k(\rho_n) \sum_{i_1, \dots, i_k > 0} \delta^j(i_1 + \dots + i_k) \frac{\nu_n(i_1)}{\rho_n} \dots \frac{\nu_n(i_k)}{\rho_n}. \end{aligned}$$

Next we compute the expected number of clusters of size $j > 0$ involved in a given merger. On one hand, this should be exactly the expected number of clusters

involved in a merger, times $\nu_n(j)$. On the other, this can be computed by counting the number of clusters of size j in every cluster merger involving at least one cluster with non-zero size. This gives

$$(2.12) \quad \left(\sum_{k \geq 0} k \hat{\pi}(k) \right) \nu_n(j) \\ = \sum_{k \geq 1} \hat{R}_k(\rho_n) \sum_{i_1, \dots, i_k > 0} (\delta^j(i_1) + \dots + \delta^j(i_k)) \frac{\nu_n(i_1)}{\rho_n} \dots \frac{\nu_n(i_k)}{\rho_n}.$$

Using the criticality assumption (2.6), subtracting (2.12) from (2.11) and observing that the $k = 1$ term cancels, we obtain (2.7). \square

Remark. Without the criticality assumption (2.6), the above shows that (2.7) should be replaced with

$$(2.13) \quad \nu_{n+1}(j) - \nu_n(j) = (\Xi - 1)\nu_n(j) + \sum_{k \geq 2} \hat{R}_k(\rho_n) \hat{I}_k(\nu_n, j), \quad j > 0,$$

if

$$(2.14) \quad \Xi \stackrel{\text{def}}{=} \sum_{k \geq 0} k \hat{\pi}(k),$$

the expected number of clusters involved in a simultaneous merger, is finite.

3. GALTON-WATSON PROCESSES AND COAGULATION DYNAMICS.

The purpose of this section is to establish a direct and elementary connection between critical Galton-Watson processes and the coalescence processes introduced in Section 2.

3.1. A brief introduction to Galton-Watson processes. The Galton-Watson 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, and is replaced by his 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

$$P_n(i, j) \stackrel{\text{def}}{=} \mathbf{P}\{X_n = j \mid X_0 = i\}$$

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 Galton-Watson process is to choose an array $\{\xi_{i,j}\}$ of independent, identically distributed random variables with distribution $\hat{\pi}$, and

define

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

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

Remark. The formula (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 Galton-Watson 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.

Proposition 3.1. *Let X be a Galton-Watson process, and let*

$$(3.2) \quad \hat{\nu}_n(j) \stackrel{\text{def}}{=} \mathbf{P}\{X_n = j\} = P_n(1, j).$$

denote the distribution of X_n . Suppose 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,

$$\hat{\nu}_n = \nu_n$$

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 (2.7)–(2.10).

Recall a Galton-Watson 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 (2.6), and it is well known that such processes become extinct almost surely.

Proof of Proposition 3.1. According to the branching formula (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

$$(3.3) \quad \hat{\nu}_{n+1}(j) = \sum_{i=0}^{\infty} \hat{\nu}_n(i) \hat{\pi}^{*i}(j) = \sum_{i=0}^{\infty} P_n(1, i) P_1(i, j).$$

However, one also has another characterization, due to the Markov property. Namely,

$$(3.4) \quad \hat{\nu}_{n+1}(j) = \sum_{k \geq 0} P_1(1, k) P_n(k, j) = \sum_{k \geq 0} \hat{\pi}(k) \hat{\nu}_n^{*k}(j),$$

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

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 [21, 22]. 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 (2.7)–(2.10) is expected.

4. BERNSTEIN TRANSFORM OF THE DISCRETE EVOLUTION EQUATIONS

The beautiful, classical theory of Galton-Watson 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 n^{th} -generation descendant distribution (ν_n) is then given by the n^{th} functional iterate of G —As is well-known and is easy to derive from (2.3), the function

$$(4.1) \quad G_n(z) = \sum_{j=0}^{\infty} \nu_n(j)z^j$$

satisfies $G_0(z) = z$ and

$$(4.2) \quad G_{n+1}(z) = G(G_n(z)), \quad n \geq 0.$$

In order to simplify the study of continuum limits and 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 Galton-Watson process X , we define its *Bernstein transform* by

$$(4.3) \quad \hat{\varphi}_n(q) = \mathbf{E}(1 - e^{-qX_n}) = \sum_{j \geq 1} \nu_n(j)(1 - e^{-qj}).$$

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

$$(4.4) \quad \hat{\varphi}_n(q) = 1 - G_n(e^{-q}).$$

The function $\hat{\varphi}_n$ is a *Bernstein function* [36]—a nonnegative function whose derivative is a Laplace transform. The class of Bernstein functions has a number of convenient properties—e.g., it is closed under composition and taking pointwise limits—and such functions have proved valuable in previous work on coagulation dynamics (see for example [28, 30, 31]). The first objective of this section is to obtain a convenient expression (equation (4.5), below) for the evolution of the Bernstein transform of a Galton-Watson process.

Proposition 4.2. *If $\hat{\varphi}$ is the Bernstein transform of a Galton-Watson process X , then*

$$(4.5) \quad \hat{\varphi}_{n+1}(q) - \hat{\varphi}_n(q) = -\hat{\Psi}(\hat{\varphi}_n(q)) \quad \text{for all } q \geq 0,$$

where

$$(4.6) \quad \hat{\Psi}(s) \stackrel{\text{def}}{=} \sum_{j=0}^{\infty} (1-s)^j \hat{\pi}(j) - 1 + s = G(1-s) - 1 + s,$$

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

Proof. By (4.4) and (4.2) we have

$$(4.7) \quad \hat{\varphi}_{n+1}(q) - \hat{\varphi}_n(q) = 1 - G(G_n(e^{-q})) - \hat{\varphi}_n(q) = -\hat{\Psi}(\hat{\varphi}_n(q)). \quad \square$$

Definition 4.3. We define the function $\hat{\Psi}$ in equation (4.6) to be the *branching mechanism* of the Galton-Watson process X .

While the generating function has proved extremely useful in many contexts, the notion of branching mechanism as defined above is better suited for our purposes. As will be seen, it is strongly analogous to the branching mechanism of a CSBP, it governs convergence of the process in the continuum limit (Proposition 6.3, below), and it also provides the discrete analog of the Poissonian structure of the rate constants arising in [12] (see (4.9) below, 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.

Lemma 4.4. *If the expected family size $\Xi = \sum_{k \geq 0} k \hat{\pi}(k)$ is finite, then the branching mechanism of the Galton-Watson process X satisfies*

$$(4.8) \quad \hat{\Psi}(s) = \sum_{j=2}^{\infty} ((1-s)^j - 1 + js) \hat{\pi}(j) + (1 - \Xi)s.$$

We now compute the rate constants \hat{R}_k (defined in (2.9)) in terms of the branching mechanism.

Lemma 4.5. *Let X be a Galton-Watson process with branching mechanism $\hat{\Psi}$. Then for all $k \geq 2$ we have*

$$(4.9) \quad \hat{R}_k(\rho_n) = \frac{(-\rho_n)^k}{k!} \hat{\Psi}^{(k)}(\rho_n),$$

where ρ_n is defined by (2.8) and \hat{R}_k is defined by (2.9).

Proof. Termwise differentiation of (4.6) gives

$$(4.10) \quad \frac{(-1)^k}{k!} \hat{\Psi}^{(k)}(s) = \sum_{j \geq k} \hat{\pi}(j) \binom{j}{k} (1-s)^{j-k}, \quad k \geq 2,$$

for $s \in (0, 2)$. Because $\rho_n \in (0, 1)$ for all n , we may substitute $s = \rho_n$ in (4.10) and this gives (4.9) for $k \geq 2$. \square

Part 2. Convergence criteria governing scaling limits of Galton-Watson processes.

5. CSBPs AND THE TIME CONTINUOUS SMOLUCHOWSKI EQUATION.

In subsequent sections we will consider scaling limits of Galton-Watson processes. 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 [12] 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 [14, 26] or [23, Chap. 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 non-decreasing sample paths and formally solves $\partial_t \Theta = x + \bar{X}_\Theta$ with $\Theta(x, 0) = 0$.² In this context, the Laplace exponent of \bar{X}_t , denoted Ψ , is called the *branching mechanism* for $Z_t(x)$ and has Lévy-Khintchine representation

$$(5.1) \quad \Psi(q) = \frac{1}{2}a_0q^2 - aq - b + \int_{(0, \infty)} (e^{-qx} - 1 + qx \mathbb{1}_{\{x < 1\}}) d\pi(x),$$

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

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

It is well known [17, 23] that a CSBP remains finite almost surely (or is *conservative*) if and only if

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

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

$$(5.2) \quad 0 \leq \Psi'(0^+) = -a - \int_{[1, \infty)} x d\pi(x).$$

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

$$(5.3) \quad \Psi(q) = \frac{1}{2}a_0q^2 + a_\infty q + \int_{(0, \infty)} (e^{-qx} - 1 + qx) d\pi(x),$$

for some $a_\infty \geq 0$, together with the finiteness condition

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

Note $a_\infty = \Psi'(0^+)$ and can be expressed in terms of a using (5.2).

We remark that (5.3) closely parallels the form of the discrete branching mechanism $\hat{\Psi}$ in Definition 4.3. Indeed, for critical Galton-Watson processes $\hat{\Psi}$ takes the form (4.8), which is obtained from (5.3) by dropping the quadratic term $\frac{1}{2}a_0q^2$, setting $a_\infty = 0$ by the criticality condition, and using the approximation $e^{-s} \approx 1 - s$.

²A derivation of an analog of this for Galton-Watson processes is in Remark 7.2, below.

We will see later (Proposition 6.3) 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

$$(5.4) \quad \mathbf{E}(e^{-qZ_t(x)}) = e^{-x\varphi_t(q)},$$

where the spatial Laplace exponent φ solves the backward equation

$$(5.5) \quad \partial_t \varphi_t(q) = -\Psi(\varphi_t(q)), \quad \text{with initial data } \varphi_0(q) = q.$$

This is exactly the continuum analog of the discrete equation (4.5), and we will provide a natural and elementary proof of it in Proposition 6.4, below.

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

$$(5.6) \quad \varphi_t(q) = b_t q + \int_{(0,\infty)} (1 - e^{-qx}) d\bar{\nu}_t(x), \quad q \geq 0,$$

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

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

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 [12] shows that the Lévy jump measure of a *critical* CSBP which becomes extinct almost surely satisfies a generalized type of *Smoluchowski coagulation equation*. Explicitly, Proposition 3 in [12] shows that

$$(5.7) \quad \partial_t \langle f, \bar{\nu}_t \rangle = \sum_{k \geq 2} R_k I_k(\bar{\nu}_t, f), \quad \text{for all } f \in C([0, \infty]),$$

where

$$(5.8) \quad I_k(\bar{\nu}, f) \stackrel{\text{def}}{=} \int_{(0,\infty)^k} \left(f(x_1 + \dots + x_k) - \sum_{i=1}^k f(x_i) \right) \prod_{i=1}^k \frac{d\bar{\nu}(x_i)}{\langle 1, \bar{\nu} \rangle},$$

$$(5.9) \quad R_k \stackrel{\text{def}}{=} \frac{(-\rho_t)^k \Psi^{(k)}(\rho_t)}{k!}, \quad \rho_t \stackrel{\text{def}}{=} \langle 1, \bar{\nu}_t \rangle.$$

Here

$$\langle f, \bar{\nu}_t \rangle \stackrel{\text{def}}{=} \int_{(0,\infty)} f(x) d\bar{\nu}_t(x),$$

is the duality pairing.

Equation (5.7) has a natural interpretation as a coagulation model introduced by Smoluchowski [40, 41] 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

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

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

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

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 (5.7) models coalescence.

In general the rate constants R_k appearing in (5.7) can be chosen arbitrarily. In the context of CSBPs, the R_k 's have a special Poissonian structure given by (5.9). 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 limit of the n^{th} generation descendant distributions of rescaled critical Galton-Watson processes ν_n . To briefly explain the main idea, recall Proposition 3.1 shows that ν_n satisfies (2.7). This is of course simply a discrete version of (5.7). Indeed, given a sequence $\{f(j)\}_{j \in \mathbb{N}}$, we multiply (2.7) by $f(j)$ and sum over j to obtain

$$(5.10) \quad \sum_{j>0} f(j)(\nu_{n+1}(j) - \nu_n(j)) = \sum_{k \geq 2} \hat{R}_k \hat{I}_k(\nu_n, f),$$

where

$$(5.11) \quad \hat{I}_k(\nu_n, f) \stackrel{\text{def}}{=} \sum_{i_1, \dots, i_k > 0} \left[f\left(\sum_{l=1}^k i_l\right) - \sum_{l=1}^k f(i_l) \right] \prod_{l=1}^k \frac{\nu_n(i_l)}{\rho_n}.$$

Moreover, Lemma 4.5 shows that the rate constants are obtained from the discrete branching mechanism $\hat{\Psi}$ (Definition 4.3) in exactly the same manner as (5.9).

Thus, after rescaling the Galton-Watson processes correctly, it is only natural to expect convergence of the rescaled branching mechanisms of the Galton-Watson 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 (5.7) if the limit is critical, and (5.7) with an additional damping term (analogous to (2.13)) if the limit is subcritical. We prove this in Proposition 6.3 and Corollaries 6.5 and 6.8 below.

6. SCALING LIMITS OF CRITICAL GALTON-WATSON PROCESSES.

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

We establish necessary and sufficient criteria for convergence of the discrete branching mechanisms (Proposition 6.3), convergence of the Bernstein transforms (Proposition 6.4) and of the rescaled Galton-Watson processes themselves (Proposition 6.6), in terms of a type of weak convergence of the reproduction laws alone. Moreover, we show (Corollaries 6.5 and 6.8) 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 (2.8)–(2.10) (where \hat{R}_k is defined by (2.9)). 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 $h, \tau \rightarrow 0$, we associate measures supported on the grid $h\mathbb{N}$ to the rescaled size distributions. Explicitly, we define

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

Here δ_x denotes the Dirac measure centered at x .

In the context of Galton-Watson processes, the above corresponds to scaling 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 Galton-Watson 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

$$(6.2) \quad \varphi_n^h(q) = \int_{\mathbb{R}^+} (1 - e^{-qx}) d\nu_n^h(x) = \frac{1}{h} \hat{\varphi}_n(hq).$$

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

$$(6.3) \quad \Psi_{h,\tau}(q) \stackrel{\text{def}}{=} \int_{\mathbb{R}^+} ((1 - hq)^{x/h} - 1 + qx) d\pi_{h,\tau}(x) = \frac{\hat{\Psi}(hq)}{\tau h}.$$

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 $0 \leq q \leq 1/h$ 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), equation (5.10) becomes*

$$(6.4) \quad \frac{\langle f, \nu_{n+1}^h \rangle - \langle f, \nu_n^h \rangle}{\tau} = \sum_{k \geq 2} R_k^{h,\tau}(\rho_n^h) I_k(\nu_n^h, f)$$

for any bounded $f \in C(\mathbb{R}_+)$. Here I_k is defined by (5.8), and

$$(6.5) \quad R_k^{h,\tau}(\rho_n^h) \stackrel{\text{def}}{=} \frac{(-\rho_n^h)^k}{k!} \Psi_{h,\tau}^{(k)}(\rho_n^h),$$

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

$$(6.6) \quad \boxed{\frac{\varphi_{n+1}^h(q) - \varphi_n^h(q)}{\tau} = -\Psi_{h,\tau}(\varphi_n^h(q))} \quad \text{for all } n \geq 0, q \geq 0.$$

Proof. The proof is a direct computation using (5.10), Lemma 4.5, and (4.5). \square

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 the Appendix.

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 satisfy the criticality condition (2.6), and let (h_k) , (τ_k) be positive sequences converging to 0. We introduce (rescaled) discrete branching mechanisms as in (4.6) and (6.3) with $\hat{\pi}_k$ replacing $\hat{\pi}$, by defining

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

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

$$(6.7) \quad d\hat{\mu}_k(x) \stackrel{\text{def}}{=} \sum_{j \geq 2} (j-1) \hat{\pi}_k(j) d\delta_j(x).$$

This measure is rescaled according to the relation

$$(6.8) \quad d\check{\mu}_k(x) \stackrel{\text{def}}{=} \frac{1}{\tau_k} d\hat{\mu}_k\left(\frac{x}{h_k}\right) = \frac{1}{\tau_k} \sum_{j \geq 2} (j-1) \hat{\pi}_k(j) d\delta_{jh_k}(x).$$

The next definition singles 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.5).

Definition 6.2. Given some finite measure κ on $[0, \infty]$, we say the sequence $(\check{\mu}_k)$ Lévy-converges to κ provided

$$(6.9) \quad (x \wedge 1) d\check{\mu}_k(x) \rightarrow d\kappa(x) \quad \text{weak-}\star \text{ on } [0, \infty].$$

Recall 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 $\langle g, \kappa_k \rangle \rightarrow \langle g, \kappa \rangle$. We require test functions to be continuous at ∞ in order to capture any atom at ∞ .

Proposition 6.3. *Given a sequence $(\hat{\pi}_k)$ satisfying the criticality condition (2.6), and positive sequences (h_k) , (τ_k) converging to zero, let $(\check{\mu}_k)$, $(\check{\Psi}_k)$ be as above.*

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

$$(6.10) \quad \check{\Psi}_k(q) \rightarrow \Psi(q) \stackrel{\text{def}}{=} \frac{1}{2} \alpha_0 q^2 + \alpha_\infty q + \int_0^\infty \frac{e^{-qx} - 1 + qx}{x} d\mu(x),$$

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

$$(6.11) \quad d\kappa(x) = \alpha_0 d\delta_0 + \alpha_\infty d\delta_\infty + (x \wedge 1) d\mu(x).$$

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

(ii) Conversely, suppose $\Psi(q) = \lim_{k \rightarrow \infty} \check{\Psi}_k(q)$ exists for each $q \in [0, \infty)$. Then $(\check{\mu}_k)$ Lévy-converges to some finite measure κ on $[0, \infty]$, and Ψ is given by (6.10)–(6.11).

Remark. The Lévy-convergence requirement (6.9) corresponds exactly to convergence of Lévy triples in a natural topology associated to subordinators. Explicitly, criterion (6.9) is equivalent to convergence of the Lévy triples $(0, 0, \check{\mu}_k)$ to the Lévy triple $(\alpha_0, \alpha_\infty, \mu)$, as described in the Appendix.

Remark. Recall that the expression for Ψ in (6.10) is the general form of a branching mechanism for a finite (sub)critical CSBP (see [23, Ch. 12] or Section 5). We show in Section 8, below, that every such branching mechanism does arise as a sequential limit from discrete branching mechanisms of *critical* Galton-Watson processes. A heuristic explanation as to why critical branching mechanisms might yield a subcritical branching mechanism in the limit is discussed in Remark 6.7, below.

Proof of Proposition 6.3. We begin by proving (i).

1. Define

$$(6.12) \quad d\kappa_k = (x \wedge 1) d\check{\mu}_k.$$

Then $\kappa_k \rightarrow \kappa$ weak- \star on $[0, \infty]$ by (6.9). Next fix $q \geq 0$, and compute

$$(6.13) \quad \Psi'(q) = \alpha_0 q + \alpha_\infty + \int_0^\infty (1 - e^{-qx}) d\mu(x) = \langle f_0, \kappa \rangle,$$

$$(6.14) \quad \check{\Psi}'_k(q) = \int_{(h, \infty)} \frac{1 - (1 - h_k q)^{(x-h_k)/h_k}}{x - h_k} x d\check{\mu}_k(x) = \langle f_{h_k}, \kappa_k \rangle,$$

where

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

The second equality in (6.14) follows because κ_k is supported on $[2h_k, \infty)$.

Note $\Psi(0) = \Psi_{h_k}(0) = 0$, and Ψ' and Ψ'_{h_k} are positive and increasing. Thus, the desired conclusion in (6.10) for fixed q will follow, provided we show

$$(6.15) \quad \langle f_{h_k}, \kappa_k \rangle = \langle f_{h_k} - f_0, \kappa_k \rangle + \langle f_0, \kappa_k \rangle \rightarrow \langle f_0, \kappa \rangle.$$

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

$$(6.16) \quad \limsup_{h \rightarrow 0} \sup_{x > h} |f_h(x) - f_0(x)| = 0.$$

Because $\langle 1, \kappa_k \rangle \rightarrow \langle 1, \kappa \rangle$, this immediately implies (6.15).

2. To finish the proof of (6.10), it only remains to prove (6.16). For this, we claim that, provided $\max(h, qh) < \frac{1}{2}$,

$$(6.17) \quad |f_h(x) - f_0(x)| \leq \begin{cases} 5e^{-qx} + \frac{2h}{x} & \text{for } x > 2, \\ 2(1 \vee x)hq^2 & \text{for } x > h. \end{cases}$$

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

$$0 < \frac{x}{x-h} - 1 = \frac{h}{x-h} < \frac{2h}{x},$$

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

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

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

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

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

By adding these last two bounds and using $x/(x \wedge 1) = 1 \vee x$, we infer (6.17).

Using the first estimate in (6.17) for $x > h^{-1/2}$, and the second estimate in (6.17) for $x \leq h^{-1/2}$, we obtain (6.16). This finishes the proof of (6.10).

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, because of (6.14), the functions $(\check{\Psi}'_k)_{k \geq N}$ 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 finishes the proof of part (i).

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

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

Therefore by (6.14) we have that whenever k is so large that $2qh_k < 1$,

$$(6.18) \quad (\inf f_0) \langle 1, \kappa_k \rangle \leq \langle f_{h_k}, \kappa_k \rangle = \check{\Psi}'_k(q) \leq \frac{\check{\Psi}_k(2q)}{q}.$$

The last inequality holds because $\check{\Psi}_k$ is convex and positive on $(q, 2q)$. Because $\inf f_0 > 0$ and $(\check{\Psi}_k(2q))$ 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 part (i) above, we infer that for any $q > 0$, as $k \rightarrow \infty$ along the appropriate subsequence,

$$\check{\Psi}'_k(q) = \langle f_{h_k}, \kappa_k \rangle \rightarrow \Phi(q) \stackrel{\text{def}}{=} \langle f_0, \kappa \rangle.$$

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

Let $(\alpha_0, \alpha_\infty, \mu)$ be the Lévy triple associated with κ by (6.11). Then we see that (6.13) holds, and Ψ' is the Bernstein transform of this triple. Because the Bernstein transform is bijective (see Theorem A.3 of the Appendix), it follows that Ψ determines κ uniquely. By consequence, the whole sequence (κ_k) must converge, meaning that (6.9) holds. This finishes the proof. \square

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 (6.6) as a forward Euler difference approximation to the ODE (5.5).

Proposition 6.4. *Make the same assumptions as in Proposition 6.3.*

(i) *Suppose that the Lévy convergence in (6.9) holds for some finite measure κ on $[0, \infty]$. Then, as $k \rightarrow \infty$ and whenever $n\tau_k \rightarrow t$, we have*

$$(6.19) \quad \varphi_n^{h_k}(q) \rightarrow \varphi(q, t) \quad \text{for every } q, t \in [0, \infty),$$

where φ is the unique solution of

$$(6.20) \quad \partial_t \varphi(q, t) = -\Psi(\varphi(q, t)), \quad \varphi(q, 0) = q,$$

with Ψ given by (6.10). Moreover, $\varphi(\cdot, t)$ is Bernstein for each $t \geq 0$, and has the form

$$(6.21) \quad \varphi(q, t) = \beta_0(t)q + \beta_\infty + \int_{(0, \infty)} (1 - e^{-qx}) d\nu_t(x),$$

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

(ii) *Conversely, suppose (6.19) holds, and $\varphi(q_0, t_0) > 0$ for some $q_0, t_0 > 0$. Then (6.9) holds for some finite measure κ on $[0, \infty]$.*

Proof. We prove (i) by providing a rather straightforward proof that the forward Euler difference scheme in (6.6) converges if (6.10) holds. Fix $q \in \mathbb{R}^+$ and $T > 0$. From equation (6.20) we see

$$\varphi(q, t) = q - \int_0^t \Psi(\varphi(q, s)) ds.$$

Because $s \mapsto \Psi(\varphi(q, s))$ 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

$$(6.22) \quad 0 \leq R'_1 \leq \tau_k \sum_{m=0}^{n-1} \left(\Psi(\varphi(q, m\tau_k)) - \Psi(\varphi(q, (m+1)\tau_k)) \right) \leq \tau_k \Psi(q).$$

Let $\varphi_k(q, n\tau_k) \stackrel{\text{def}}{=} \varphi_n^{h_k}(q)$ and $\Psi_k \stackrel{\text{def}}{=} \Psi_{h_k, \tau_k}$. Summing (6.6), and noting $\nu_0^h = \delta_h/h$, we have

$$(6.23) \quad \begin{aligned} \varphi_k(q, n\tau_k) &= \frac{1 - e^{-qh_k}}{h_k} - \tau_k \sum_{m=0}^{n-1} \Psi_k(\varphi_k(q, m\tau_k)) \\ &= q - \tau_k \sum_{m=0}^{n-1} \Psi(\varphi_k(q, m\tau_k)) + R'_2, \end{aligned}$$

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

$$(6.24) \quad |R'_2| \leq M_k \stackrel{\text{def}}{=} q^2 h_k + T \sup_{[0, q]} |\Psi_k - \Psi|.$$

Consequently, because 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,

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

Proposition 6.3 guarantees $M_k \rightarrow 0$ as $k \rightarrow \infty$. Because $\varphi(q, t)$ has bounded derivative $|\partial_t \varphi| \leq \Psi(q)$, we may infer that whenever $n\tau_k \rightarrow t \in [0, T]$,

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

This finishes the proof of convergence of (φ_k) . Because $\varphi_k(\cdot, n\tau_k)$ is Bernstein, and any pointwise limit of Bernstein functions is Bernstein [36, Cor. 3.7], the limit $\varphi(\cdot, t)$ is Bernstein and has the form in (6.21) for some Lévy triple $(\beta_0, \beta_\infty, \nu_t)$. We must have $\beta_\infty = 0$, however, because $\varphi(q, t) \leq q$ for all $q, t \geq 0$.

For the proof of (ii), choose $q, t > 0$ such that $\varphi(q, t) > 0$. For k sufficiently large, 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 we find from (6.23) that

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

Hence $\{\Psi_k(\varphi(q, t)/2)\}$ is bounded. Following the proof of Proposition 6.3 (specifically, using (6.18)) we see that the set of measures $\{\kappa_k\}$ is weak- \star pre-compact. (Here κ_k is defined by (6.12).)

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 part (i) above, we know (6.20) holds with Ψ given by taking the limit in (6.10) along the appropriate subsequence, as in the proof of part (ii) of Proposition 6.3. Then φ is C^1 , and Ψ is determined by φ by evaluating (6.20) at $t = 0$. As in the proof of Proposition 6.3, it follows κ is determined by φ , hence the whole sequence (κ_k) converges. This means (6.9) holds, finishing the proof. \square

Because convergence of the Bernstein transforms is equivalent to convergence of corresponding Lévy triples and weak- \star convergence of corresponding measures on $[0, \infty]$ by the continuity theorem A.6, we obtain the following corollary.

Corollary 6.5. *Under the same assumptions as Proposition 6.4 part (i), we have that for every $t \geq 0$, as $k \rightarrow \infty$ and whenever $nh_k \rightarrow t$,*

$$(6.25) \quad (x \wedge 1) d\nu_n^{h_k}(x) \rightarrow \beta_0(t) d\delta_0(x) + (x \wedge 1) d\nu_t(x) \quad \text{weak-}\star \text{ on } [0, \infty].$$

Conversely, if this convergence holds, then (6.9) 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 (6.10) we have $\Psi'(\infty) = \infty$, or equivalently

$$(6.26) \quad a_0 > 0 \quad \text{or} \quad \int_{(0, \infty)} d\mu(x) = \infty.$$

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

Moreover, we have $\langle 1, \nu_t \rangle < \infty$ if and only if Grey's condition holds:

$$(6.27) \quad \int_{[1, \infty)} \frac{du}{\Psi(u)} < \infty.$$

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

$$(6.28) \quad \partial_t \langle f, \bar{\nu}_t \rangle = -\Psi'(0^+) \langle f, \bar{\nu}_t \rangle + \sum_{k \geq 2} R_k I_k(\bar{\nu}_t, f),$$

for all test functions $f \in C([0, \infty])$. Here R_k and I_k are as in (5.8) and (5.9) respectively.

Remark. In the critical case $\Psi'(0^+) = 0$ and (6.28) is precisely the generalized Smoluchowski equation (5.7) encountered before.

Remark. The convergence property (6.25) explains precisely how the Lévy jump measure of a (sub)critical CSBP X arises as a limit of scaled n -th generation descendant distributions. This was one of our main motivations for developing this treatment of continuum limits of Galton-Watson processes.

Remark. For the case when Grey's condition (6.27) does not hold, Bertoin and Le Gall have proposed an interesting generalization of Smoluchowski's equation in terms of the sum of locations of atoms of Poisson random measures; see [12, Eq. (26)].

Proof. 1. The weak convergence in (6.25) and the converse follow immediately from Proposition 6.3 and the continuity theorem for Bernstein transforms (Theorem A.6).

2. Next, note that by (6.21) and because $(1 - e^{-qx})/1 \leq x \wedge (1/q) \rightarrow 0$ as $q \rightarrow \infty$,

$$(6.29) \quad \beta_0(t) = \lim_{q \rightarrow \infty} \frac{\varphi(q, t)}{q}.$$

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

$$q \geq \int_0^t \Psi(\varphi(q, s)) ds \geq t\Psi(\varepsilon q)$$

for all $q > 0$. Because Ψ is convex it follows $\Psi'(q) \leq 1/(\varepsilon t)$ for all q . Conversely, if $\Psi'(q) \leq M$ for all q then $\Psi(q) \leq Mq$ and $\varphi(q, t) \geq qe^{-Mt}$ by (6.20).

Now, the condition $\lim_{q \rightarrow \infty} \Psi'(q) = \infty$ is easily seen to be equivalent to (6.26). This establishes both as necessary and sufficient to have $\beta_0(t) \equiv 0$. In the case $\Psi'(\infty) < \infty$, then due to (6.20) and (6.29) we find

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

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

3. That Grey's condition is necessary and sufficient for $\langle 1, \nu_t \rangle < \infty$ is well-known (see for example [23, proof of Theorem 12.5]) and is easily deduced by separating variables in (6.20) and integrating.

4. Finally, we show that (ν_t) satisfies (6.28). In the case that Ψ is critical (i.e. $\Psi'(0^+) = 0$) a proof can be found in [12, Proposition 3] or [20, Theorem 3.2]. In the

general case the proof closely resembles [20, Theorem 3.2] and we sketch the details here.

The main idea is to establish (6.28) for test functions of the form $f_q(x) \stackrel{\text{def}}{=} 1 - e^{-qx}$. Indeed, observe

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

Consequently

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

where second equality is true because 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 (6.20) and the fact that $\Psi(0) = 0$.

This proves (6.28) 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 Theorem 3.2 in [20]. \square

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

$$(6.30) \quad Y_t^{(k)}(x) = h_k X_{\lfloor t/\tau_k \rfloor, k} \left(\left\lfloor \frac{x}{h_k} \right\rfloor \right).$$

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

Proposition 6.6. *Make the same assumptions as in Proposition 6.3.*

(i) *Suppose there exists a finite measure κ on $[0, \infty]$ such that $(\check{\mu}_k)$ Lévy-converges (as in Definition 6.2) to κ . Then the finite-dimensional distributions of the rescaled Galton-Watson processes $Y^{(k)}$ converge to those of a finite (sub)critical CSBP Z . Further, the Bernstein transforms of $Y^{(k)}$'s converge pointwise to the Laplace exponent of Z . That is, if*

$$\begin{aligned} \varphi_k(q, t) &\stackrel{\text{def}}{=} \frac{1}{h_k} \mathbf{E} [1 - \exp(-qY_t^{(k)}(h_k))], \\ \text{and} \quad \varphi(q, t) &\stackrel{\text{def}}{=} -\ln \mathbf{E} [\exp(-qZ_t(1))] \end{aligned}$$

then

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

for every $q \in \mathbb{R}^+$ and $t \geq 0$.

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

Remark 6.7. Before presenting the proof, we momentarily pause to remark on criticality. Each process $Y^{(k)}$ is a critical Galton-Watson process, however, the limit need not be critical. Indeed, the branching mechanism of the limit is given by (6.10) 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 $\check{\mu}_k$ Lévy-converge to a measure that has an atom at ∞ . Physically this corresponds to the situation when the rescaled descendant distributions $\check{\mu}_k$ 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. First assume (6.9) holds. By Proposition 6.4 we know that φ_k converges to a function φ which satisfies (6.20). We know from [20] (see also [12]) that the solution of (6.20) is the Laplace exponent of a (sub)critical CSBP Z . To finish 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

$$\begin{aligned} \mathbf{E}\left(\exp(-qY_t^{(k)}(x))\right) &= \mathbf{E}\left(\exp(-qY_t^{(k)}(N_k h_k))\right) \\ &= \left[\mathbf{E}\left(\exp(-qY_t^{(k)}(h_k))\right)\right]^{N_k} = \left[1 - h_k \varphi_k(q, t)\right]^{N_k} \\ &\xrightarrow{k \rightarrow \infty} e^{-\varphi(q, t)x} = \mathbf{E}\left(\exp(-qZ_t(x))\right) \end{aligned}$$

This shows 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 [27, page 280].

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

$$\mathbf{E}\left(\exp(-qY_{t_0}^{(k)}(x))\right) \xrightarrow{k \rightarrow \infty} e^{-\varphi(q, t_0)x} = \mathbf{E}\left(\exp(-qZ_{t_0}(x))\right).$$

Thus

$$\lim_{k \rightarrow \infty} \left[1 - h_k \varphi_k(q, t_0)\right]^{N_k} \text{ exists and equals } e^{-\varphi(q, t_0)x},$$

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

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

By our assumption on Z , we know $\varphi(q, t_0) > 0$ and we can apply Proposition 6.4. This will guarantee (6.9) holds. By using part (i), we infer there is a finite (sub)critical CSBP with the same one-dimensional distributions as Z . This completes the proof. \square

As an immediate corollary, we prove a special case of a result in [12] showing that the Lévy jump measure of certain CSBPs satisfy the Smoluchowski equation.

Corollary 6.8. *Let Z be a finite (sub)critical CSBP whose branching mechanism satisfies Grey's condition (6.27), and ν_t be the Lévy jump measure associated with Z_t . Then ν is a weak solution of the damped Smoluchowski equations (6.28).*

Proof. We note first that there exists a sequence of processes $Y^{(k)}$ of the form (6.30) such that the finite-dimensional distributions of $Y^{(k)}$ converge to Z . While the existence of such a sequence is well known [18, 29] we will prove a much stronger result in Theorem 8.1, below. By Proposition 6.6 we know that the Bernstein transforms of $Y^{(k)}$ converge pointwise to the Laplace exponent of Z . Consequently, by Corollary 6.5 the convergence (6.25) holds, where ν_t is the Lévy jump measure of Z_t , and further ν satisfies (6.28) as desired. \square

7. GRIMVALL'S CONVERGENCE CRITERION, SIMPLIFIED AND COMPARED.

Grimvall [18] established a criterion that is necessary and sufficient for convergence of rescaled Galton-Watson processes in general, in terms of convolution powers of family-size distributions, shifted and scaled. Here, using the classical theory of infinite divisibility from [15], we describe how Grimvall's criterion can be expressed more simply in terms of weak convergence of canonical measures. Then we show that for critical Galton-Watson processes, Grimvall's result is equivalent to Proposition 6.6.

7.1. A simplification of Grimvall's result. We begin by stating Grimvall's result using notation consistent with ours.

Theorem 7.1 (Grimvall [18], Theorems 3.1 and 3.3). *Let (h_k) , (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*

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

- (1) *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 (6.30) converge to those of a (possibly infinite) CSBP with an absorbing state $+\infty$.*
- (2) *Conversely, if for every $t \in [0, 1]$ the random variables $Y_t^{(k)}$ converge weakly to a random variable Y_t , with $\mathbf{P}\{Y_1 > 0\} > 0$, then $\eta_k^{*1/(h_k\tau_k)}$ converges weakly to a probability measure η .*

Remark 7.2. 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

$$\Theta_0 = 0, \quad \Theta_{n+1} \stackrel{\text{def}}{=} \Theta_n + x + \bar{X}_{\Theta_n}.$$

The Lamperti transformation defines a process X by

$$X_n \stackrel{\text{def}}{=} x + \bar{X}_{\Theta_n}.$$

Observe

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

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 Galton-Watson process 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.

Our aim in this subsection is to show that Grimvall's criterion can be expressed in terms of simple weak convergence criteria on $[0, \infty)$ that do not involve convolution powers of η_k . Toward this end, we introduce the (overloaded) notation

$$(7.1) \quad d\hat{\pi}_k(x) = \sum_{j \geq 0} \hat{\pi}_k(j) d\delta_j(x)$$

to denote the probability measures on \mathbb{R} determined by the distributions $\hat{\pi}_k$ on \mathbb{N}_0 . Then $d\eta_k(x) = d\hat{\pi}_k((x + h_k)/h_k)$. We recall that a sequence of finite measures (λ_k) on an interval $I \subset \mathbb{R}$ converges *narrowly* to λ on I if $\langle f, \lambda_k \rangle \rightarrow \langle f, \lambda \rangle$ for every $f \in C_b(I)$, the space of bounded continuous functions on I .

Proposition 7.3. *The sequence $(\eta_k^{*1/(h_k\tau_k)})$ converges weakly to a probability measure η on \mathbb{R} if and only if, as $k \rightarrow \infty$,*

$$(7.2) \quad b_k \stackrel{\text{def}}{=} \int_{[0, \infty)} \sin(x - h_k) \frac{d\hat{\pi}_k(x/h_k)}{\tau_k h_k} \rightarrow b,$$

and

$$(7.3) \quad ((x - h_k)^2 \wedge 1) \frac{d\hat{\pi}_k(x/h_k)}{\tau_k h_k} \rightarrow K$$

narrowly on $[0, \infty)$, where $b \in \mathbb{R}$ and K is a finite measure supported on $[0, \infty)$, such that

$$(7.4) \quad \int_{\mathbb{R}} e^{izx} d\eta(x) = \exp \left(ibz + \int_{\mathbb{R}} \frac{e^{izx} - 1 - iz \sin x}{x^2 \wedge 1} dK(x) \right)$$

is the Lévy-Khintchine representation of the infinitely divisible distribution η .

Remark. We emphasize that neither Theorem 7.1 nor Proposition 7.3 require the criticality assumption (2.6). While (2.6) is implicitly assumed throughout the rest of this paper, it is not required in the following proof.

Proof. We know [15, §XVII.1, Thm 1] that $(\eta_k^{*1/(h_k\tau_k)})$ converges weakly to a probability measure η if and only if

$$\lim_{k \rightarrow \infty} \frac{1}{h_k\tau_k} \int_{\mathbb{R}} (e^{ix\xi} - 1) d\eta_k(x)$$

exists for all $\xi \in \mathbb{R}$ and is continuous in ξ . This is equivalent (see [15, §XVII.2, Lemma 1 and Thm. 2]) to the existence of $b \in \mathbb{R}$ and a canonical measure M such

that, as $k \rightarrow \infty$,

$$(7.5) \quad \int_{\mathbb{R}} \sin x \frac{d\eta_k(x)}{h_k \tau_k} \rightarrow b,$$

$$(7.6) \quad (x^2 \wedge 1) \frac{d\eta_k(x)}{h_k \tau_k} \rightarrow \left(1 \wedge \frac{1}{x^2}\right) dM(x),$$

narrowly on \mathbb{R} . We recall that a canonical measure M is a positive measure on \mathbb{R} for which the measure K defined by $dK(x) = (1 \wedge x^{-2}) dM(x)$ is finite, and note that Feller's definition of convergence of canonical measures $M_n \rightarrow M$ is equivalent to the convergence $(1 \wedge x^{-2})M_n(dx) \rightarrow K$ narrowly on \mathbb{R} .

Now, clearly

$$\int_{\mathbb{R}} \sin x \frac{d\eta_k(x)}{h_k \tau_k} = \int_{[0, \infty)} \sin(x - h_k) \frac{d\hat{\pi}_k(x/h_k)}{\tau_k h_k} = b_k,$$

so (7.5) is equivalent to (7.2). Next, suppose (7.5) and (7.6) hold. Let $f \in C_b(\mathbb{R})$ be a bounded continuous test function and observe

$$(7.7) \quad \int_{\mathbb{R}} f(x) (x^2 \wedge 1) \frac{d\eta_k(x)}{h_k \tau_k} = \int_{[0, \infty)} f(x - h_k) ((x - h_k)^2 \wedge 1) \frac{d\hat{\pi}_k(x/h_k)}{\tau_k h_k}.$$

We need to show that translating f by h_k does not affect the limit of the right hand side. Choose $\varepsilon > 0$ arbitrary and find $R > 2$ so that $K(R, \infty) < \varepsilon$. Without loss of generality we may assume R is a point of continuity for the measure K . Since f is uniformly continuous on $[0, R]$, for k large enough we can arrange $|f(x) - f(x - h_k)| < \varepsilon$ for all $x \in [0, R]$.

Writing

$$d\lambda_k(x) = ((x - h_k)^2 \wedge 1) d\hat{\pi}_k(x/h_k) / (\tau_k h_k),$$

we have

$$\begin{aligned} \int_{[0, \infty)} |f(x) - f(x - h_k)| d\lambda_k(x) &\leq \int_{[0, R]} \varepsilon d\lambda_k(x) + 2\|f\|_{L^\infty} \int_{(R, \infty)} d\lambda_k(x) \\ &\xrightarrow{k \rightarrow \infty} \varepsilon K[0, R] + 2\|f\|_{L^\infty} K(R, \infty) < C\varepsilon. \end{aligned}$$

Thus, on the right hand side of (7.7), the test function f can be translated by h_k , which immediately implies (7.3).

The converse follows using a similar argument. \square

Remark. In our proof we showed that test functions can be translated by h_k when tested against the measure $d\lambda_k(x)$. The weight $(x - h_k)^2 \wedge 1$ however can not be translated similarly!

7.2. Equivalence in the critical case. The main result of this section shows that under the criticality assumption (2.6), the conditions (7.2) and (7.3) are equivalent to the Lévy-convergence (Definition 6.2) of the rescaled Lévy measures $\check{\mu}_k$ determined from family size distributions $\hat{\pi}_k$ as in (6.8). What makes this a curious point is that Lévy-convergence occurs only on the compactified positive half-line $[0, \infty]$, but the convolution powers of the measures η_k in Theorem 7.1 have support on the negative half-line also. This complicated Grimvall's analysis of their convergence.

Proposition 7.4. *Let $(\hat{\pi}_k)$, be a sequence of measures satisfying the criticality assumption (2.6), (h_k) , (τ_k) be two positive sequences that converge to 0, and let $\check{\mu}_k$ be defined by (6.8). Then (7.2) and (7.3) are equivalent to the existence of a finite measure κ on $[0, \infty]$ such that $(\check{\mu}_k)$ Lévy-converges to κ .*

Proof. 1. Suppose first (7.2) and (7.3) hold, and define $\check{\mu}_k$ so that

$$(7.8) \quad d\check{\mu}_k(x) = \frac{1}{\tau_k h_k} \sum_{j \geq 2} (j h_k - h_k) \hat{\pi}_k(j) d\delta_{j h_k} = (x - h_k) d\check{\pi}_k + \frac{\hat{\pi}_k(0)}{\tau_k} d\delta_0,$$

where for brevity we define

$$(7.9) \quad d\check{\pi}_k \stackrel{\text{def}}{=} \frac{d\hat{\pi}_k(x/h_k)}{\tau_k h_k}.$$

Our first step is to show that the measures $(x \wedge 1) d\check{\mu}_k$ are uniformly bounded. We start by proving the tail bound

$$(7.10) \quad \sup_k \int_{(1, \infty)} d\check{\mu}_k(x) < \infty.$$

To see this we use criticality (2.6) to infer $\int_{[0, \infty)} (x - h_k) d\check{\pi}_k = 0$ and hence

$$\int_{(1, \infty)} d\check{\mu}_k(x) = \int_{(1, \infty)} (x - h_k) d\check{\pi}_k(x) = - \int_{[0, 1]} (x - h_k) d\check{\pi}_k(x).$$

Because $|x - \sin x| \leq x^2$ for $|x| \leq 1$ and $|\sin x| \leq 1$, we then have

$$\left| \int_{(1, \infty)} d\check{\mu}_k(x) - b_k \right| \leq \int_{[0, \infty)} ((x - h)^2 \wedge 1) d\check{\pi}_k(x),$$

whence (7.10) holds, by (7.2) and (7.3).

Next we claim

$$(7.11) \quad \sup_k \int_{[0, 1]} x d\check{\mu}_k(x) = \sup_k \int_{[0, 1]} x(x - h_k) d\check{\pi}_k(x) < \infty.$$

Indeed, choosing the test function $f(x) = 1$, equation (7.3) and criticality give

$$\begin{aligned} K[0, \infty) &= \lim_{k \rightarrow \infty} \int_{[0, \infty)} ((x - h_k)^2 \wedge 1 + h_k(x - h_k)) d\check{\pi}_k(x) \\ &= \lim_{k \rightarrow \infty} \left(\int_{[0, 1+h_k]} x d\check{\mu}_k(x) + \int_{(1+h_k, \infty)} (1 + h_k(x - h_k)) d\check{\pi}_k(x) \right). \end{aligned}$$

Combined with (7.10) this gives (7.11) as claimed.

2. The bounds (7.10) and (7.11) show that $\{(x \wedge 1) d\check{\mu}_k\}$ is weak- \star precompact in the set of finite measures on $[0, \infty]$. Now choose any subsequence (which we again index by k for convenience) along which

$$(x \wedge 1) d\check{\mu}_k(x) \rightarrow \kappa = a_0 d\delta_0 + a_\infty d\delta_\infty + (x \wedge 1) d\mu(x)$$

weak- \star on $[0, \infty]$. We claim a_0 , a_∞ and μ can directly be determined from K and b via the identities

$$(7.12) \quad (x \wedge 1) d\mu(x) = (1 \vee x) dK(x) \quad \text{on } \mathbb{R}^+,$$

$$(7.13) \quad a_0 = K\{0\},$$

$$(7.14) \quad a_\infty + \int_{(0, \infty)} \left(1 - \frac{\sin x}{x}\right) d\mu(x) = -b.$$

This will show that subsequential limits are unique, proving (6.9). Thus, to complete the proof, it remains to establish (7.12)–(7.14).

i. To prove (7.12), let $\varepsilon > 0$ and let f be a continuous function supported in $[\varepsilon, 1/\varepsilon]$. Observe that due to (7.3) and the uniform continuity of $x/(x^2 \wedge 1)$ on $[\varepsilon/2, 2/\varepsilon]$,

$$\begin{aligned} \int_{(0,\infty)} f(x) d\kappa(x) &= \lim_{k \rightarrow \infty} \int_{\mathbb{R}^+} f(x)(x \wedge 1) d\check{\mu}_k(x) \\ &= \lim_{k \rightarrow \infty} \int_{\mathbb{R}^+} f(x)(x \wedge 1) \frac{(x - h_k)}{(x - h_k)^2 \wedge 1} ((x - h_k)^2 \wedge 1) d\check{\pi}_k(x) \\ &= \lim_{k \rightarrow \infty} \int_{\mathbb{R}^+} f(x)(x \wedge 1) \frac{x}{x^2 \wedge 1} ((x - h_k)^2 \wedge 1) d\check{\pi}_k(x) \\ &= \int_{\mathbb{R}^+} f(x)(1 \vee x) dK(x). \end{aligned}$$

This proves (7.12). Moreover, incidentally it follows $\int_{(1,\infty)} x dK(x) < \infty$.

ii. To prove (7.13), choose $f(x) = (1 \vee x)^{-1} = 1 \wedge (1/x)$ and observe that, because (7.10) ensures

$$\lim_{k \rightarrow \infty} h_k \int_{(1,\infty)} (x - h_k) d\check{\pi}_k(x) = 0 = \lim_{k \rightarrow \infty} h_k \int_{(1,\infty)} x^{-1} d\check{\pi}_k(x),$$

and furthermore $|1 - (x - h)| \leq 2h$ for $x \in [1, 1 + h]$, we have

$$\begin{aligned} a_0 + K(\mathbb{R}^+) &= \langle f, \kappa \rangle = \lim_{k \rightarrow \infty} \int_{\mathbb{R}^+} f(x)(x \wedge 1) (x - h_k) d\check{\pi}_k(x) \\ &= \lim_{k \rightarrow \infty} \left(\int_{(0,1]} x(x - h_k) d\check{\pi}_k(x) + \int_{(1,\infty)} \left(\frac{x - h_k}{x} \right) d\check{\pi}_k(x) \right) \\ &= \lim_{k \rightarrow \infty} \int_{\mathbb{R}^+} ((x - h_k)^2 \wedge 1) d\check{\pi}_k(x) = K[0, \infty). \end{aligned}$$

iii. Finally, to prove (7.14) define

$$f(x) = \frac{1}{x \wedge 1} \left(1 - \frac{\sin x}{x} \right), \quad f(0) = 0, \quad f(\infty) = 1,$$

and observe

$$\begin{aligned} a_\infty + \int_{\mathbb{R}^+} \left(1 - \frac{\sin x}{x} \right) d\mu(x) &= \lim_{k \rightarrow \infty} \int_{\mathbb{R}^+} f(x)(x \wedge 1) d\check{\mu}_k(x) \\ &= \lim_{k \rightarrow \infty} \int_{[0,\infty)} \left(1 - \frac{\sin x}{x} \right) (x - h_k) d\check{\pi}_k(x) \\ (7.15) \quad &= -b + \lim_{k \rightarrow \infty} \int_{[0,\infty)} \left(\frac{\sin(x - h_k)}{x - h_k} - \frac{\sin x}{x} \right) (x - h_k) d\check{\pi}_k(x), \end{aligned}$$

where we use criticality and (7.2) for the last equality. Due to (7.10), we have

$$\lim_{k \rightarrow \infty} \int_{(1,\infty)} \left(\frac{\sin(x - h_k)}{x - h_k} - \frac{\sin x}{x} \right) (x - h_k) d\check{\pi}_k(x) = 0.$$

Now, because $|(d/dx)(\sin x/x)| \leq |x|$, whenever $h_k \leq |x - h_k| \leq 1$ we have $|x| \leq 2|x - h_k|$ and

$$\left| \frac{\sin(x - h_k)}{x - h_k} - \frac{\sin x}{x} \right| \leq 2h_k |x - h_k|.$$

Because indeed $h_k \leq |x - h_k|$ on the support of $(x - h_k) d\check{\pi}_k$, then, due to (7.3) the second term on the right of (7.15) converges to 0. This proves (7.14), and completes the proof of the forward implication.

3. Conversely, assume (6.9). Using criticality and the fact that $h_k \leq x - h_k$ on the support of $\check{\mu}_k$, we immediately see by integrating (7.8) that $h_k^2 \hat{\pi}_k(0)/(\tau_k h_k)$ is bounded, and we deduce that $((x - h_k)^2 \wedge 1) d\check{\pi}_k$ and b_k , the left hand sides of (7.2) and (7.3), are bounded, as in step 1 above. Now an argument similar to the above shows that for any subsequential limits K and b , the identities (7.12)–(7.14) hold. This shows uniqueness of subsequential limits and implies (7.2) and (7.3), concluding the proof. \square

Part 3. Universality in Galton-Watson process and CSBPs.

8. UNIVERSAL GALTON-WATSON FAMILY SIZE DISTRIBUTIONS

The following result is motivated by the existence of universal eternal solutions of Smoluchowski's coagulation equations [31], which in turn was inspired by Feller's account of *Doebelin's universal laws* in classical probability theory [15, Section XVII.9].

Theorem 8.1. *There exists a Galton-Watson process X with family-size distribution $\hat{\pi}: \mathbb{N} \rightarrow [0, \infty)$ and sequences $(h_n), (\tau_n) \rightarrow 0$, with the following property: For any (sub)critical branching mechanism Ψ taking the form in (6.10), there is a subsequence along which the rescaled process $Y^{(k)}$ (defined by rescaling X in equation (6.30)) 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 Section 7 in [31]. The terminology used in [31], however, is substantially different from our terminology which is based on [36] and Appendix A, below.³ Thus for the readers convenience we provide a self-contained treatment here.

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 right notions of convergence and scaling of Lévy triples.

The right notion of convergence of Lévy triples is a generalization of Lévy-convergence as introduced earlier in Definition 6.2. Namely, to each Lévy triple $\lambda \stackrel{\text{def}}{=} (\alpha_0, \alpha_\infty, \mu)$ we associate κ , a finite measure on $[0, \infty]$, by

$$(8.1) \quad d\kappa(x) = a_0 d\delta_0(x) + a_\infty d\delta_\infty(x) + (x \wedge 1) d\mu(x).$$

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

Definition 8.2. Let (λ_k) be a sequence of Lévy triples, and κ_k the associated measures defined as in (8.1) above. We say (λ_k) converges if the sequence (κ_k) converges weak- \star in the spaces of finite measures on $[0, \infty]$.

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

³The correspondence in terminology is described in Remark A.8, below

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

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

where

$$(8.2) \quad \alpha_0^{b,c} = cb^{-1} \alpha_0, \quad \alpha_\infty^{b,c} = c \alpha_\infty, \quad \text{and} \quad d\mu^{b,c}(x) = c d\mu(bx).$$

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

Definition 8.3. Let λ_\star be a Lévy triple and $(b_k), (c_k)$ be two sequences that converge to infinity. We say λ_\star is a *universal Lévy triple* with sequences $(b_k), (c_k)$ if for any Lévy triple λ we have

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

Our next lemma shows that this notion of universality is completely determined by the tail of λ_\star .

Lemma 8.4. 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 ν_\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 $(b_k), (c_k)$.

Remark. Lemma 8.4 is still true if finitely many terms of the sequences $(b_k), (c_k)$ are arbitrarily changed.

To prove Lemma 8.4 it helps to introduce left and right distribution functions as follows.

Definition 8.5. 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

$$\kappa_L(x) \stackrel{\text{def}}{=} \alpha_0 + \int_{(0,x]} z d\mu(z), \quad \kappa_R(x) \stackrel{\text{def}}{=} \alpha_\infty + \int_{(x,\infty)} d\mu.$$

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 the appendix (Theorem A.7). We can now prove Lemma 8.4.

Proof of Lemma 8.4. 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 $(c_k/b_k) \rightarrow 0$. Once this is established, the lemma immediately follows from the fact that

$$(c_k b_k^{-1} \alpha_0, 0, |\mu_\star^{b_k, c_k} - \nu_\star^{b_k, c_k}|) \rightarrow 0$$

in the topology of Lévy triples.

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

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

Now by monotonicity of $\kappa_{\star L}$ and universality of λ_{\star} it follows that $\kappa_{\star L}(y) \rightarrow \infty$ as $y \rightarrow \infty$. Moreover, (8.3) and the continuity theorem (Theorem A.7) imply that $\kappa_{\star L}^{b_k, c_k}(x)$ converges (along the chosen subsequence) to a finite limit for some x . This forces $(c_k/b_k) \rightarrow 0$ along the chosen subsequence, finishing the proof. \square

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

Proposition 8.6. *There exist sequences (b_k) , (c_k) such that the set of Lévy triples of the form $(0, 0, \mu_{\star})$ which are universal with respect to (b_k) , (c_k) is dense in the space of all Lévy triples.*

We postpone the proof of Proposition 8.6 until the proof of Theorem 8.1 is complete.

Proof of Theorem 8.1. Using Proposition 8.6 we choose a Lévy triple $\lambda_{\star} = (0, 0, \mu_{\star})$ that is universal with sequences (b_k) , (c_k) . Using Lemma 8.4 we can, without loss of generality, assume

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

Let

$$(8.5) \quad \tau_k = c_k^{-1}, \quad h_k = b_k^{-1},$$

and define a family size distribution $\hat{\pi}(j)$, $j \geq 0$, by

$$(8.6) \quad \hat{\pi}(0) = 1 - \int_{\mathbb{R}^+} d\mu_{\star}(x),$$

$$(8.7) \quad \hat{\pi}(j) = \frac{1}{j-1} \int_{(j-2, j-1]} d\mu_{\star}(x), \quad j \geq 2,$$

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

$$(8.8) \quad d\hat{\mu}(x) = \sum_{j \geq 2} (j-1) \hat{\pi}(j) d\delta_j(x) = \sum_{j \geq 2} \left(\int_{(j-2, j-1]} d\mu_{\star}(x) \right) d\delta_j(x).$$

We claim that the Lévy triple $\hat{\lambda} \stackrel{\text{def}}{=} (0, 0, \hat{\mu})$ is also universal. Once universality of $\hat{\lambda}$ is established, subsequential convergence of $Y^{(k)}$ follows immediately from Proposition 6.6.

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

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

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

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

$$(8.9) \quad \frac{h_k}{\tau_k} \kappa_{\star L} \left(\frac{x}{h_k} \right) \rightarrow \kappa_L(x), \quad \frac{1}{\tau_k} \kappa_{\star R} \left(\frac{x}{h_k} \right) \rightarrow \kappa_R(x),$$

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

$$(8.10) \quad \frac{h_k}{\tau_k} \hat{\kappa}_L \left(\frac{x}{h_k} \right) \rightarrow \kappa_L(x), \quad \frac{1}{\tau_k} \hat{\kappa}_R \left(\frac{x}{h_k} \right) \rightarrow \kappa_R(x),$$

where $(\hat{\kappa}_L, \hat{\kappa}_R)$ are the pair of left and right distribution functions associated to $\hat{\lambda}$.

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

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

which means

$$\kappa_{\star R}(z) \leq \hat{\kappa}_R(z) \leq \kappa_{\star R}(z-2).$$

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

$$(8.11) \quad \kappa_R(x) \leq \liminf \frac{1}{\tau_k} \hat{\kappa}_R \left(\frac{x}{h_k} \right) \leq \limsup \frac{1}{\tau_k} \hat{\kappa}_R \left(\frac{x}{h_k} \right) \leq \kappa_R(x_-).$$

Thus the second limit in (8.10) follows.

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

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

and this entails

$$\kappa_{\star L}(z-1) \leq \hat{\kappa}_L(z) \leq \kappa_{\star L}(z) + 2.$$

Using the argument in Lemma 8.4 we see $b_k/c_k = h_k/\tau_k \rightarrow 0$. Consequently, for any two points of continuity $x_- < x$ for κ_L , as above we infer

$$(8.12) \quad \kappa_L(x_-) \leq \liminf \frac{h_k}{\tau_k} \hat{\kappa}_L \left(\frac{x}{h_k} \right) \leq \limsup \frac{h_k}{\tau_k} \left(\hat{\kappa}_L \left(\frac{x}{h_k} \right) + 2 \right) \leq \kappa_L(x).$$

The second limit in (8.10) follows, and this proves universality of the coarse grained measure μ_\star . This finishes the existence part of Theorem 8.1.

To prove density, we use Lemma 8.4 to note that any probability measure on \mathbb{N}_0 with the same tail at $\hat{\pi}$ is also universal. Since all such probability measures are dense in the space of all probability measures on \mathbb{N}_0 , we obtain density. This finishes the proof of Theorem 8.1. \square

It remains to prove Proposition 8.6. Fix a sequence $c_k \rightarrow \infty$ that satisfies

$$(8.13) \quad \sum_{j=1}^{\infty} \frac{j}{c_j} < 1, \quad c_k \sum_{j>k} \frac{j}{c_j} \rightarrow 0 \quad \text{as } k \rightarrow \infty.$$

For example, $c_k = ce^{k^2}$ works for small enough $c > 0$. The main tool used in the proof of Proposition 8.6 is the following ‘‘packing lemma’’, which is the analog of Lemma 7.2 in [31].

Lemma 8.7 (Packing lemma). *Let $\lambda_k = (0, 0, \mu_k)$ be a sequence of Lévy triples, let Φ_k be the corresponding Bernstein transforms given by*

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

and assume that

$$(8.14) \quad \int_{\mathbb{R}^+} d\mu_k(x) \leq k, \quad \text{for all } k \in \mathbb{N}.$$

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

$$(8.15) \quad \Phi_\star(q) \stackrel{\text{def}}{=} \sum_{j=1}^{\infty} c_j^{-1} \Phi_j(b_j q)$$

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

$$(8.16) \quad c_k \Phi_\star(b_k^{-1} q) - \Phi_k(q) \rightarrow 0 \quad \text{as } k \rightarrow \infty, \quad \text{for all } q > 0.$$

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

Proof. As $\Phi_j(b_j q) \leq \Phi_j(\infty) \leq j$ for all $q > 0$, the series (8.15) is uniformly bounded and converges for each q . We estimate the quantity in (8.16) in two parts, to show how b_k can be chosen. Note

$$\left| \Phi_k(q) - c_k \sum_{j=k+1}^{\infty} c_j^{-1} \Phi_j(b_k^{-1} b_j q) \right| \leq c_k \sum_{j=k+1}^{\infty} j c_j^{-1} \rightarrow 0$$

as $k \rightarrow \infty$, regardless of what b_k is. Then because $\Phi_j(q) \rightarrow 0$ as $q \rightarrow 0$ 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 (8.16) follows.

The limit function Φ_\star is the Bernstein transform of some Lévy triple $\lambda_\star = (a_0, a_\infty, \mu_\star)$ by Theorem A.6. Because $\Phi_j(0^+) = 0$, we infer

$$\Phi_\star(0^+) \leq \sum_{j=k+1}^{\infty} j c_j^{-1}$$

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

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

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

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 Lemma 7.3 in [31].

Lemma 8.8. *Let $\lambda = (a_0, a_\infty, \mu)$ be a Lévy triple. Then there is a sequence of measures μ_k that satisfy (8.14), such that the triples*

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

Proof. For the proof it suffices to consider λ of the form $\lambda = (0, 0, \mu)$, because these are dense in the space of Lévy triples. But then there exist $\varepsilon_k \rightarrow 0$ such that $\mu_k \stackrel{\text{def}}{=} \mathbb{1}_{\{x \geq \varepsilon_k\}} \mu$ satisfies $\int_{\mathbb{R}^+} d\mu_k \leq \frac{k}{3}$. Evidently, $\lambda_k = (0, 0, \mu_k) \rightarrow \lambda$ due to Definition 8.2. \square

With this we can prove Proposition 8.6.

Proof of Proposition 8.6. Because the space of finite measures on $[0, \infty]$ with the weak- \star topology is separable, the same is true 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 8.8, select measures μ_k satisfying (8.14) such that $\lambda_k = (0, 0, \mu_k) \rightarrow \bar{\lambda}_n$ along the n^{th} subsequence. Construct a sequence b_k , a Bernstein function Φ_\star and its associated Lévy triple λ_\star using Lemma 8.7. We claim λ_\star is universal with sequences $(b_k), (c_k)$.

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

$$(8.17) \quad \Phi_\star^{b_k, c_k}(q) = c_k \Phi_\star\left(\frac{q}{b_k}\right).$$

Now given any Lévy triple λ , choose a subsequence (indexed by k) along which $(\bar{\lambda}_k) \rightarrow \lambda$. If $\bar{\Phi}_k$ is the Bernstein transform of $\bar{\lambda}_k$, then the continuity theorem for Bernstein transforms (Theorem A.6) guarantees $(\bar{\Phi}_k) \rightarrow \Phi$ pointwise. By (8.16) this implies $(\Phi_\star^{b_k, c_k}) \rightarrow \Phi$, and by the continuity theorem again we have $(\lambda_\star^{b_k, c_k}) \rightarrow \lambda$ along this subsequence. This finishes the proof. \square

9. LINEARIZATION AND UNIVERSALITY FOR CRITICAL CSBP

The dilational form of the scaling relations (8.2), (8.4) and (8.17) hint at exact scaling relations that hold for the limiting CSBPs, which we develop in this section. These relations establish that the *nonlinear dynamics of CSBPs becomes 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 9.3 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 [31] that showed that scaling dynamics on the scaling attractor for solvable Smoluchowski equations becomes linear and dilational in terms of a Lévy-Khintchine representation. The definition of the scaling attractor in [31] as the collection of all limits of rescaled solutions was motivated by 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) becomes 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

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

If the branching mechanism Ψ is (sub)critical, and given by (6.10) in terms of 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

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

and this corresponds to left and right distribution functions given by the scaling relations

$$(9.3) \quad \tilde{\kappa}_L(x) = cb^{-1}\kappa_L(bx), \quad \tilde{\kappa}_R(x) = c\kappa_R(bx).$$

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

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

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

and the second relation follows by considering the ODE (5.5) 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 (6.10), and using the definitions in (A.7). \square

For the study of long-time scaling limits of CSBPs, we should take b, c as functions of t with $c(t) \rightarrow \infty$ or as sequences $(b_k), (c_k)$ with $c_k \rightarrow \infty$. 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 6.4 and 6.6 for scaling limits of Galton-Watson processes.

For brevity in expressing our result, let us say that the Lévy triple $\lambda = (\alpha_0, \alpha_\infty, \mu)$ in (6.10) *generates* the branching mechanism Ψ and the corresponding (finite, subcritical) CSBP Z .

Theorem 9.3 (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 8.2. Then the finite-dimensional distributions of Z_k converge to those of the finite (sub)critical CSBP Z generated by λ .*

(ii) *Conversely, suppose the one-dimensional distributions of Z_k converge to those of some finite process Z such that $\mathbf{P}\{Z_{t_0}(x_0) > 0\} > 0$ 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. To prove (i), suppose λ_k converges to λ . Then, due to the second continuity theorem A.7, the branching mechanisms Ψ_k generated by λ_k converge pointwise to the branching mechanism Ψ generated by λ . Due to the representation formula (A.6) and an argument using Montel's theorem analogously to the proof of Proposition 6.3, the convergence occurs locally uniformly for every derivative. Then it is straightforward to show that because the Laplace exponents φ_k of Z_k satisfy

$$(9.4) \quad \partial_t \varphi_k(q, t) = -\Psi_k(\varphi_k(q, t)), \quad \varphi_k(q, 0) = q,$$

one has $\varphi_k(q, t) \rightarrow \varphi(q, t)$ as $k \rightarrow \infty$ for each $q > 0$, $t \geq 0$, where φ satisfies (6.20). 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 [27, page 280].

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

$$e^{-x\varphi_k(q, t)} = \mathbf{E}(e^{-qZ_k(x, t)}) \rightarrow \mathbf{E}(e^{-qZ(x, t)})$$

for each $x, t > 0$, we have that

$$\varphi(q, t) \stackrel{\text{def}}{=} \lim_{k \rightarrow \infty} \varphi_k(q, t)$$

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 because Ψ_k is increasing and convex, for large enough k we find $2q < \varphi_k(\hat{q}, t_0)$, hence by integrating (9.4) we find

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

Hence with κ_k determined from λ_k by (A.4) and $g_q(x)$ defined by (A.5), we find

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

As in the proof of the continuity theorem A.6 (converse part), this implies that $\langle 1, \kappa_k \rangle$ is bounded. Hence the sequence of Lévy measures (λ_k) is precompact. 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 λ is unique, and the whole sequence (λ_k) converges. \square

9.2. Existence of universal critical CSBPs. The continuity theorem 9.3 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 the present paper. A large number of results exist in the classical literature that cover supercritical and subcritical cases; see [2, 17, 23, 25, 38, 39], e.g., for further details. For relatively recent results on critical cases we refer to [20, 32, 33]. Our paper [20] provides necessary and sufficient criteria for approach to self-similar form for critical CSBP that become extinct almost surely (having branching mechanism satisfying Grey's condition), under a quite general assumption on the scaling that is taken, but assuming there is a unique limit as $t \rightarrow \infty$.

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 Galton-Watson family size distributions in Section 8. This observation gives a precise meaning to a remark made by Grey [17] 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 9.4. *There exists a finite critical CSBP Z_\star that is universal in the following sense: There exist sequences $(b_k), (c_k) \rightarrow \infty$ 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*

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

Proof of Theorem 9.4. Using Proposition 8.6 choose $\lambda_\star = (0, 0, \mu_\star)$ to be a universal Lévy triple with sequences $b_k, c_k \rightarrow \infty$. Let Z_\star be the critical CSBP generated by λ_\star . The theorem now follows immediately from the continuity theorem (Theorem 9.3). \square

Remark. Let φ_\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

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

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

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

and a similar expression for $\tilde{\varphi}$, cf. (6.21). We may note that

$$\beta_{\star 0}(t) = \exp(-\Psi'_\star(\infty)t), \quad \Psi'_\star(\infty) = \int_0^\infty d\mu_\star(x) < \infty,$$

and that $\beta_{\star 0}(c_k t) \rightarrow 0$ as $k \rightarrow \infty$. Due to the continuity theorem A.6, the convergence above corresponds to a Lévy convergence property: for each $t > 0$, as $k \rightarrow \infty$,

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

APPENDIX A. CONTINUITY THEOREMS FOR BERNSTEIN TRANSFORMS OF LÉVY TRIPLES.

The Lévy-convergence requirement (Definition 6.2) that is used in all our results relates to a natural topology of Lévy triples associated with subordinators. The purpose of 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.

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

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

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

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

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

The recent book of Schilling et al. [36] develops the theory of Bernstein functions extensively. The main representation theorem regarding these functions (see Theorem 3.2 of [36]) 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.3. *A function $f: \mathbb{R}^+ \rightarrow \mathbb{R}$ is Bernstein if and only if it has the representation*

$$(A.2) \quad f(q) = a_0 q + a_\infty + \int_{\mathbb{R}^+} (1 - e^{-qx}) d\mu(x),$$

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

For convenience, we call the function f in (A.2) 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 instance [7]) that the Laplace exponent of a subordinator can be uniquely expressed in the form (A.2), 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

$$(A.3) \quad d\kappa(x) = a_0 d\delta_0(x) + a_\infty d\delta_\infty(x) + (x \wedge 1) d\mu(x).$$

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

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

Definition A.4. We say 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

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

Remark A.5. This is exactly the same as Definition 8.2, restated here for convenience. Moreover, this generalizes the notion of Lévy-convergence introduced Definition 6.2. Indeed, Lévy-convergence of $(\check{\mu}_k)$ to a measure κ is exactly convergence of the Lévy triples $(0, 0, \check{\mu}_k)$ to the associated Lévy triple $(\kappa(0), \kappa(\infty), \kappa|_{(0, \infty)})$.

The following provides a continuity theorem for the Bernstein transform that is not present in [36]. It may be inferred from the proof of Theorem 3.1 in [31], but [31] employs a different and rather cumbersome terminology, and the proof below is much simpler.

Theorem A.6. *Let $(a_0^{(k)}, a_\infty^{(k)}, \mu_k)$ be a sequence of Lévy triples with Bernstein transforms f_k corresponding via (A.2) and measures κ_k corresponding via (A.3). Then the following are equivalent.*

- (i) $f(q) \stackrel{\text{def}}{=} \lim_{k \rightarrow \infty} f_k(q)$ exists for each $q \in (0, \infty)$.
(ii) κ_k converges to some finite measure κ weak- \star on $[0, \infty]$. That is, as $k \rightarrow \infty$,
- $$\langle g, \kappa_k \rangle \rightarrow \langle g, \kappa \rangle \quad \text{for every } g \in C([0, \infty]).$$

If either condition holds, then the limits f, κ correspond to a unique Lévy triple via (A.2) and (A.3) 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.4.

Proof. Let (κ_k) be the sequence of measures given by

$$(A.4) \quad d\kappa_k(x) = a_0^{(k)} d\delta_0(x) + a_\infty^{(k)} d\delta_\infty(x) + (x \wedge 1) d\mu_k(x).$$

Suppose first that $\kappa_k \rightarrow \kappa$ weak- \star on $[0, \infty]$. For $q \in \mathbb{R}^+$, consider the test function $g_q : \mathbb{R}^+ \rightarrow \mathbb{R}$ defined by

$$(A.5) \quad g_q(x) = \frac{1 - e^{-qx}}{x \wedge 1}.$$

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

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

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

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

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

Consequently,

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

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.3). Because 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 . \square

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: (a) The *Bernstein primitive* (or branching mechanism)

$$(A.6) \quad \psi(q) = \frac{1}{2} a_0 q^2 + a_\infty q + \int_{\mathbb{R}^+} \frac{e^{-qx} - 1 + qx}{x} d\mu(x),$$

and (b) the pair of *left and right distribution functions* (κ_L, κ_R) given by

$$(A.7) \quad \kappa_L(x) = a_0 + \int_{(0, x]} z d\mu(z), \quad \kappa_R(x) = a_\infty + \int_{(x, \infty)} d\mu(z).$$

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.7. *Make the same assumptions as in Theorem A.6, and for each k associate ψ_k and $(\kappa_{L,k}, \kappa_{R,k})$ via (A.6) and (A.7) respectively. Then the following conditions are also equivalent to (i) and (ii).*

- (iii) $\psi(q) \stackrel{\text{def}}{=} \lim_{k \rightarrow \infty} \psi_k(q)$ exists for each $q \in (0, \infty)$.
- (iv) For almost every $x \in (0, \infty)$, both of the following limits exist:

$$(A.8) \quad \kappa_L(x) = \lim_{k \rightarrow \infty} \kappa_{L,k}(x), \quad \kappa_R(x) = \lim_{k \rightarrow \infty} \kappa_{R,k}(x).$$

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

Proof. By Theorem A.6, it suffices to show (i) is equivalent to (iii), and (ii) is equivalent to (iv).

Suppose (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

$$\psi_k(q) = \int_0^q f_k(r) dr \rightarrow \int_0^q f(r) dr \stackrel{\text{def}}{=} \psi(q).$$

Hence (iii) holds. Conversely, suppose (iii) holds. Note that $\psi_k(2q) \geq qf_k(q)$ for all $q > 0$. For each $q > 0$, it follows $(f_k(q))$ is precompact. Because f_k is increasing and concave, each subsequence has a further subsequence that converges pointwise to some (increasing and concave) function f . Necessarily,

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

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 (ii) implies (iv). Given κ as in (ii) we may define κ_L, κ_R so that

$$(A.9) \quad d\kappa_L(x) = \frac{x d\kappa(x)}{x \wedge 1}, \quad d\kappa_R(x) = \frac{d\kappa(x)}{x \wedge 1},$$

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

$$(A.10) \quad g_L(x) = \frac{x g_0(x)}{x \wedge 1}, \quad g_R(x) = \frac{g_\infty(x)}{x \wedge 1}.$$

Then as $k \rightarrow \infty$ we have

$$(A.11) \quad \langle g_0, \kappa_{L,k} \rangle = \langle g_L, \kappa_k \rangle \rightarrow \langle g_L, \kappa \rangle \stackrel{\text{def}}{=} \langle g_0, \kappa_L \rangle,$$

$$(A.12) \quad \langle g_\infty, \kappa_{R,k} \rangle = \langle g_R, \kappa_k \rangle \rightarrow \langle g_R, \kappa \rangle \stackrel{\text{def}}{=} \langle g_\infty, \kappa_R \rangle.$$

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

Conversely, assume (iv). Fix smooth cutoff functions $v_L, v_R : [0, \infty] \rightarrow [0, 1]$ such that $v_L + v_R = 1$, $v_L = 1$ on $[0, 1]$ and $v_R = 1$ on $[2, \infty]$. Given any $g \in C([0, \infty])$ we write $g = g_L + g_R$ where $g_L = gv_L$, $g_R = gv_R$, and determine g_0, g_∞ by (A.10). Then

$$(A.13) \quad \langle g_L + g_R, \kappa_k \rangle = \langle g_0, \kappa_{L,k} \rangle + \langle g_\infty, \kappa_{R,k} \rangle \rightarrow \langle g_0, \kappa_L \rangle + \langle g_\infty, \kappa_R \rangle$$

as $k \rightarrow \infty$. Because $\lim_{k \rightarrow \infty} \langle g, \kappa_k \rangle$ exists for each g , there exists a measure κ on $[0, \infty]$ with $\kappa_k \rightarrow \kappa$ weak- \star on $[0, \infty]$. Thus (ii) holds. \square

Remark A.8. As mentioned earlier the terminology used in this Appendix differs from that used in [31]. For convenience, we describe the correspondence in terminology as follows.

Let (a_0, a_∞, μ) be a Lévy triple, associated as in the present paper with the measure κ on $[0, \infty]$ defined by (A.3) and left distribution function κ_L given by (A.7). Such a triple corresponds to what is called a \bar{g} -measure (G, g_∞) in [31], according to the relations

$$G(x) = \kappa_L(x) = a_0 + \int_{(0,x]} z d\mu(z), \quad g_\infty = a_\infty.$$

In [31], what is called a g -measure G on $[0, \infty)$ corresponds here to a Lévy triple in the same way, but with $a_\infty = g_\infty = 0$. A g -measure G is *divergent* according to [31] if and only if

$$(A.14) \quad G(0^+) = a_0 > 0 \quad \text{or} \quad \int_{(0,\infty)} x^{-1} dG(x) = \int_{(0,\infty)} d\mu(x) = \infty,$$

according to this correspondence. Convergence of a sequence of g -measures or \bar{g} -measures in [31] corresponds precisely to convergence of the corresponding Lévy triples in the present paper (see Section 3 in [31]).

REFERENCES

- [1] D. J. Aldous. Deterministic and stochastic models for coalescence (aggregation and coagulation): a review of the mean-field theory for probabilists. *Bernoulli*, 5(1):3–48, 1999.
- [2] K. B. Athreya and P. E. Ney. *Branching processes*. Dover Publications, Inc., Mineola, NY, 2004. Reprint of the 1972 original [Springer, New York; MR0373040].
- [3] N. Bacaër. *A short history of mathematical population dynamics*. Springer-Verlag London, Ltd., London, 2011.
- [4] J. Berestycki, N. Berestycki, and V. Limic. A small-time coupling between Λ -coalescents and branching processes. *Ann. Appl. Probab.*, 24(2):449–475, 2014.
- [5] J. Berestycki, N. Berestycki, and J. Schweinsberg. Small-time behavior of beta coalescents. *Ann. Inst. Henri Poincaré Probab. Stat.*, 44(2):214–238, 2008.
- [6] N. Berestycki. *Recent progress in coalescent theory*, volume 16 of *Ensaio Matemáticos [Mathematical Surveys]*. Sociedade Brasileira de Matemática, Rio de Janeiro, 2009.
- [7] J. Bertoin. Subordinators, Lévy processes with no negative jumps, and branching processes, 2000.
- [8] J. Bertoin. *Random fragmentation and coagulation processes*, volume 102 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, 2006.
- [9] J. Bertoin and J.-F. Le Gall. The Bolthausen-Sznitman coalescent and the genealogy of continuous-state branching processes. *Probab. Theory Related Fields*, 117(2):249–266, 2000.
- [10] J. Bertoin and J.-F. Le Gall. Stochastic flows associated to coalescent processes. *Probab. Theory Related Fields*, 126(2):261–288, 2003.
- [11] J. Bertoin and J.-F. Le Gall. Stochastic flows associated to coalescent processes. II. Stochastic differential equations. *Ann. Inst. H. Poincaré Probab. Statist.*, 41(3):307–333, 2005.
- [12] J. Bertoin and J.-F. Le Gall. Stochastic flows associated to coalescent processes. III. Limit theorems. *Illinois J. Math.*, 50(1-4):147–181 (electronic), 2006.
- [13] M. Birkner, J. Blath, M. Capaldo, A. Etheridge, M. Möhle, J. Schweinsberg, and A. Wakolbinger. Alpha-stable branching and beta-coalescents. *Electron. J. Probab.*, 10:no. 9, 303–325, 2005.
- [14] M. E. Caballero, A. Lambert, and G. Uribe Bravo. Proof(s) of the Lamperti representation of continuous-state branching processes. *Probab. Surv.*, 6:62–89, 2009.

- [15] W. Feller. *An introduction to probability theory and its applications. Vol. II.* Second edition. John Wiley & Sons Inc., New York, 1971.
- [16] A. Gnedin, A. Iksanov, and A. Marynych. Λ -coalescents: a survey. *J. Appl. Probab.*, 51A(Celebrating 50 Years of The Applied Probability Trust):23–40, 2014.
- [17] D. R. Grey. Asymptotic behaviour of continuous time, continuous state-space branching processes. *J. Appl. Probability*, 11:669–677, 1974.
- [18] A. Grimvall. On the convergence of sequences of branching processes. *Ann. Probability*, 2:1027–1045, 1974.
- [19] C. C. Heyde and E. Seneta. *I. J. Bienaymé. Statistical theory anticipated.* Springer-Verlag, New York-Heidelberg, 1977. Studies in the History of Mathematics and Physical Sciences, No. 3.
- [20] G. Iyer, N. Leger, and R. L. Pego. Limit theorems for Smoluchowski dynamics associated with critical continuous-state branching processes. *Ann. Appl. Probab.*, 25(2):675–713, 2015.
- [21] J. F. C. Kingman. The coalescent. *Stochastic Process. Appl.*, 13(3):235–248, 1982.
- [22] J. F. C. Kingman. On the genealogy of large populations. *J. Appl. Probab.*, (Special Vol. 19A):27–43, 1982. Essays in statistical science.
- [23] A. E. Kyprianou. *Fluctuations of Lévy processes with applications.* Universitext. Springer, Heidelberg, second edition, 2014. Introductory lectures.
- [24] A. Lambert. Coalescence times for the branching process. *Adv. in Appl. Probab.*, 35(4):1071–1089, 2003.
- [25] A. Lambert. Quasi-stationary distributions and the continuous-state branching process conditioned to be never extinct. *Electron. J. Probab.*, 12:no. 14, 420–446, 2007.
- [26] J. Lamperti. Continuous state branching processes. *Bull. Amer. Math. Soc.*, 73:382–386, 1967.
- [27] J. Lamperti. The limit of a sequence of branching processes. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete*, 7:271–288, 1967.
- [28] P. Laurençot and H. van Roessel. Absence of gelation and self-similar behavior for a coagulation-fragmentation equation. *SIAM J. Math. Anal.*, 47(3):2355–2374, 2015.
- [29] Z.-H. Li. Asymptotic behaviour of continuous time and state branching processes. *J. Austral. Math. Soc. Ser. A*, 68(1):68–84, 2000.
- [30] G. Menon and R. L. Pego. Approach to self-similarity in Smoluchowski’s coagulation equations. *Comm. Pure Appl. Math.*, 57(9):1197–1232, 2004.
- [31] G. Menon and R. L. Pego. The scaling attractor and ultimate dynamics for Smoluchowski’s coagulation equations. *J. Nonlinear Sci.*, 18(2):143–190, 2008.
- [32] A. G. Pakes. Conditional limit theorems for continuous time and state branching process. In M. Ahsanullah and G. P. Yanev, editors, *Records and Branching Processes*, pages 63–103. Nova Science Publishers, Inc., 2008.
- [33] A. G. Pakes. Critical Markov branching process limit theorems allowing infinite variance. *Adv. in Appl. Probab.*, 42(2):460–488, 2010.
- [34] J. Pitman. Coalescents with multiple collisions. *Ann. Probab.*, 27(4):1870–1902, 1999.
- [35] S. Sagitov. The general coalescent with asynchronous mergers of ancestral lines. *J. Appl. Probab.*, 36(4):1116–1125, 1999.
- [36] R. L. Schilling, R. Song, and Z. Vondraček. *Bernstein functions*, volume 37 of *de Gruyter Studies in Mathematics*. Walter de Gruyter & Co., Berlin, 2010. Theory and applications.
- [37] J. Schweinsberg. Coalescent processes obtained from supercritical Galton-Watson processes. *Stochastic Process. Appl.*, 106(1):107–139, 2003.
- [38] V. A. Vatutin and A. M. Zubkov. Branching processes. I. In *Probability theory. Mathematical statistics. Theoretical cybernetics, Vol. 23*, Itogi Nauki i Tekhniki, pages 3–67, 154. Akad. Nauk SSSR, Vsesoyuz. Inst. Nauchn. i Tekhn. Inform., Moscow, 1985.
- [39] V. A. Vatutin and A. M. Zubkov. Branching processes. II. *J. Soviet Math.*, 67(6):3407–3485, 1993. Probability theory and mathematical statistics, 1.
- [40] M. von Smoluchowski. Drei Vorträge über Diffusion, Brownsche Bewegung und Koagulation von Kolloidteilchen. *Physik. Z.*, 17:557–585, 1916.
- [41] M. von Smoluchowski. Experiments on a mathematical theory of kinetic coagulation of colloid solutions. *Zeitschrift für Physikalische Chemie–Stoichiometrie und Verwandtschaftslehre*, 92(2):129–168, 1917.
- [42] H. W. Watson and F. Galton. On the probability of the extinction of families. *The Journal of the Anthropological Institute of Great Britain and Ireland*, 4:138–144, 1875.