Outbreak size distributions in epidemics with multiple stages

Tibor Antal[1] and P. L. Krapivsky[2]

[1]*School of Mathematics, Edinburgh University, Edinburgh, EH9 3JZ, UK*
[2]*Department of Physics, Boston University, Boston, MA 02215, USA*

Multiple-type branching processes that model the spread of infectious diseases are investigated. In these stochastic processes, the disease goes through multiple stages before it eventually disappears. We mostly focus on the critical multistage Susceptible-Infected-Recovered (SIR) infection process. In the infinite population limit, we compute the outbreak size distributions and show that asymptotic results apply to more general multiple-type critical branching processes. Finally using heuristic arguments and simulations we establish scaling laws for a multistage SIR model in a finite population.

I. INTRODUCTION

A critical branching process (CBP) is the fundamental stochastic process that has innumerable applications [1–5]. Although the adjective "critical" suggests that the CBP is a peculiar branching process which should require the tuning of parameters, the CBP is arguably more important than its generic sub-critical and super-critical brethren. Indeed, the sub-critical branching process quickly leads to extinction while the super-critical branching process may result in an unlimited growth, so whenever we are seeking a framework for phenomena which are driven by branching and are maintained in a stationary state, the CBP provides the proper setting.

The simplest branching process involves duplication and death, and hence called a birth-death process. Symbolically

$$A + A$$
$$\uparrow$$
$$A \longrightarrow \emptyset$$
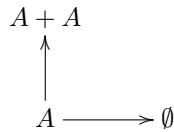
For the CBP the probabilities of both events are equal. One can think on As being cells which either divide or die. Another interpretation is to identify As with infected individuals who either die or transmit infection to other individuals. The branching process then becomes a stochastic version of the Susceptible-Infected-Recovered (SIR) infection process. According to this model [6–9], the population consists of susceptible, infected, and recovered individuals, and the infection spreads by contact between infected and susceptible members of the community, while infected individuals recover (and become immune), or equivalently they are removed from the system. The branching process also provides the simplest model of biodiversity [10] (see [11] for a survey of biological literature on macroevolution modeling), with duplication and death processes corresponding to speciations and extinctions. The CBP is *neutral* as it assumes that speciations and extinctions occur with equal probability. In the context of infectious diseases the CBPs naturally arise as a balance between human efforts leading to the reduction of the infection rate [7] and the natural evolu-

tion that increases the infection rate of diseases hovering below the threshold [12].

In numerous applications, it is desirable to generalize the CBP to include more than one type of elemental building blocks. In cell biology, we may have two populations of cells, progenitor A cells that can divide and differentiate into B cells. These two cell populations can be (on average) in a stationary state (homeostasis), if they evolve according to the two-type CBP. The simplest two-type CBP is represented by the scheme

$$A + A \qquad B + B$$
$$\uparrow \qquad \qquad \uparrow$$
$$A \longrightarrow B \longrightarrow \emptyset$$
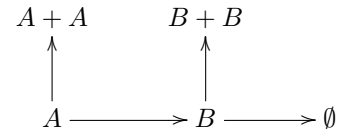
where all steps are equiprobable.

Branching processes can be studied in discrete time when all infected individuals divide or die synchronously, or in continuous time when each infected individual independently divides or dies. For a critical process the birth and death probabilities or rates are equal. Continuous-time two-type birth-death branching processes are solvable, namely the time-dependent sizes of populations of A and B individuals have been analytically determined [13]. Specializing the results of Refs. [13] to the above simplest two-type critical CBP leads to simplifications, yet the results are still cumbersome. In this paper we are concerned with different questions, namely we are interested in time-independent quantities summarizing the outbreak size distributions.

The rest of this paper is organized as follows. In Sec. II we consider the simplest multiple-type critical branching process and derive, in the infinite population limit, the outbreak size distributions. In Sec. III we show the universality of the asymptotic results for the outbreak size distributions, namely that at the epidemic threshold these distributions have algebraic tails which are determined only by the number of types (but not the detail of the model). In Sec. IV we analyze finite populations and obtain finite-size scaling laws for the sizes of the outbreaks. In Sec. V we give a short discussion. Finally, in an appendix, some of the details of the analysis of a generating function are presented.

II.   CRITICAL BIRTH-DEATH PROCESS

In this section we consider the infinite population limit. We shall always assume that the system begins with a single A, representing an infected individual. All As eventually disappear with probability one, that is the epidemic stops. We want to determine the outbreak size distribution.

A.   Single Type Critical Birth-Death Process

Let A_n be the probability that exactly n individuals catch the infection before the epidemic is over. We have $A_1 = \frac{1}{2}$. Further, $A_2 = \frac{1}{2}A_1^2$ since at the first step a new individual must get infected and then both must die without spreading infection. Proceeding along these lines we arrive at the recurrence

$$A_n = \frac{1}{2} \sum_{i+j=n} A_i A_j + \frac{1}{2}\delta_{n,1} \qquad (1)$$

reflecting that the first infection event results in two independent infection processes [2]. A solution to (1) is found by introducing the generating function

$$\mathcal{A}(z) = \sum_{n \geq 1} A_n z^n \qquad (2)$$

which converts (2) into a quadratic equation $2\mathcal{A} = \mathcal{A}^2 + z$ whose solution reads

$$\mathcal{A}(z) = 1 - \sqrt{1-z} \qquad (3)$$

Expanding $\mathcal{A}(z)$ in powers of z we find

$$A_n = \frac{1}{\sqrt{4\pi}} \frac{\Gamma\left(n - \frac{1}{2}\right)}{\Gamma(n+1)} \sim \frac{1}{\sqrt{4\pi}}\, n^{-3/2} \qquad (4)$$

In particular, the probabilities A_n are given by

$$\frac{1}{2}, \frac{1}{8}, \frac{1}{16}, \frac{5}{128}, \frac{7}{256}, \frac{21}{1024}, \frac{33}{2048}, \frac{429}{32768}, \frac{715}{65536}, \frac{2431}{262144}$$

for $n = 1, \ldots, 10$.

B.   Two-Type Critical Birth-Death Process

In this case the size distribution of the outbreaks of A types is the same as for the single type CBP. Consider now the outbreaks of B types. Denote by B_n the probability that exactly n individuals of type B are born before the outbreak is over. These probabilities satisfy

$$B_n = \frac{1}{2} \sum_{i+j=n} B_i B_j + \frac{1}{2}A_n \qquad (5)$$

To understand this recurrence it suffices to consider what happens to the initial A individual. One possibility is

that the initial A turns into a B (this happens with probability $1/2$), and then this B produces n individuals of type B with probability A_n. Combining we get the second term $\frac{1}{2}A_n$ on the right-hand side of Eq. (5). Otherwise the initial A produces two As (this happens with probability $1/2$), and then the probability that these two As will produce exactly n individuals of type B is just the convolution $\sum_{i+j=n} B_i B_j$. Overall this gives the first term on the right-hand side of Eq. (5).

To tackle Eq. (5) we again proceed by introducing the generating function

$$\mathcal{B}(z) = \sum_{n \geq 1} B_n z^n \qquad (6)$$

which converts (6) into a quadratic equation $2\mathcal{B} = \mathcal{B}^2 + \mathcal{A}$ whose solution reads

$$\mathcal{B}(z) = 1 - (1-z)^{1/4} \qquad (7)$$

Expanding $\mathcal{B}(z)$ in powers of z we find

$$B_n = \frac{1}{4\,\Gamma\left(\frac{3}{4}\right)} \frac{\Gamma\left(n - \frac{1}{4}\right)}{\Gamma(n+1)} \sim \frac{1}{4\,\Gamma\left(\frac{3}{4}\right)}\, n^{-5/4} \qquad (8)$$

C.   Multiple-Type Critical Birth-Death Process

Generally for multiple-type CBP we seemingly have

$$A_n^{(p)} = \frac{1}{2} \sum_{i+j=n} A_i^{(p)} A_j^{(p)} + \frac{1}{2} A_n^{(p-1)} \qquad (9)$$

The corresponding generating functions $\mathcal{A}^{(p-1)}$ and $\mathcal{A}^{(p)}$ are related via

$$2\mathcal{A}^{(p)} = \left[\mathcal{A}^{(p)}\right]^2 + \mathcal{A}^{(p-1)} \qquad (10)$$

We already know that

$$\mathcal{A}^{(1)} \equiv \mathcal{A} = 1 - (1-z)^{1/2}$$
$$\mathcal{A}^{(2)} \equiv \mathcal{B} = 1 - (1-z)^{1/4}$$

A general solution to (10) is

$$\mathcal{A}^{(p)} = 1 - (1-z)^{1/2^p} \qquad (11)$$

from which

$$A_n^{(p)} = \frac{1}{2^p\,\Gamma\left(1 - 2^{-p}\right)} \frac{\Gamma\left(n - 2^{-p}\right)}{\Gamma(n+1)} \sim \frac{n^{-1-2^{-p}}}{2^p \Gamma(1 - 2^{-p})} \qquad (12)$$

One can verify (11) by induction. It also follows from the more general result of the following subsection II D.

D.   Joint Outbreak Size Distribution

The joint outbreak size distribution first becomes relevant for the two-type CBP. In this setting we want to

compute the probability $P_{m,n}$ that exactly m individuals of type A and n individuals of type B are born before the outbreak is over.

The disappearance of an A results in the birth of a B, and therefore $P_{m,n} \equiv 0$ whenever $m > n$. To compute $P_{m,n} = 0$ in the interesting $m \leq n$ range, we shall use the important *abelian* property of the model: The composition of the system at the moments of birth of new Bs is irrelevant, so one can assume that the process starts with m individuals of type B (and no A individuals). We have $P_{m,m} = A_m (A_1)^m$ where the first term assures that exactly m individuals of type A ever exist and the second term gives the probability that all m individuals of type B eventually disappear without duplication. Further we have $P_{m,m+1} = A_m m (A_1)^{m-1} A_2$, and generally

$$P_{m,n} = A_m \sum_{i_1 + \ldots + i_m = n} A_{i_1} \ldots A_{i_m} \qquad (13)$$

Using the generating function

$$\mathcal{P}(x,y) = \sum_{n \geq m \geq 1} P_{m,n} x^m y^n$$

and (13) we get

$$
\begin{aligned}
\mathcal{P}(x,y) &= \sum_{m \geq 1} A_m x^m \left[\mathcal{A}(y) \right]^m \\
&= \mathcal{A}(x\mathcal{A}(y)) \\
&= 1 - \sqrt{1 - x\mathcal{A}(y)} \\
&= 1 - \sqrt{1 - x + x\sqrt{1-y}}
\end{aligned}
\qquad (14)
$$

Specializing x or y to unity we recover previous results (3) and (7):

$$\mathcal{P}(x, y=1) = \mathcal{A}(x), \quad \mathcal{P}(x=1, y) = \mathcal{B}(y) \qquad (15)$$

This essentially provides another derivation of Eq. (7).

From the generating function (14) we can express the probabilities more explicitly as

$$P_{m,n} = \frac{(2m-3)!!}{2^{m+n} n!} s_{n,n-m+1} \qquad (16)$$

where the integer $s_{n,k}$ for $n \geq k \geq 1$ are given by the recursion

$$s_{n,k} = s_{n-1,k} + (n+k-3)s_{n-1,k-1} \qquad (17)$$

with $s_{1,1} = 1$; otherwise we consider $s_{n,k}$ to be zero. The first few values of $s_{n,k}$ are as follows

$$
\begin{pmatrix}
1 & 0 & 0 & 0 & 0 & 0 \\
1 & 1 & 0 & 0 & 0 & 0 \\
1 & 3 & 3 & 0 & 0 & 0 \\
1 & 6 & 15 & 15 & 0 & 0 \\
1 & 10 & 45 & 105 & 105 & 0 \\
1 & 15 & 105 & 420 & 945 & 945
\end{pmatrix}
\qquad (18)
$$

For example, the above formulas lead to the following probabilities $P_{m,n}$ for $m, n \leq 6$

$$
\begin{pmatrix}
\frac{1}{4} & \frac{1}{16} & \frac{1}{32} & \frac{5}{256} & \frac{7}{512} & \frac{21}{2048} \\
0 & \frac{1}{32} & \frac{1}{64} & \frac{5}{512} & \frac{7}{1024} & \frac{21}{4096} \\
0 & 0 & \frac{1}{128} & \frac{3}{512} & \frac{9}{2048} & \frac{7}{2048} \\
0 & 0 & 0 & \frac{5}{2048} & \frac{5}{2048} & \frac{35}{16384} \\
0 & 0 & 0 & 0 & \frac{7}{8192} & \frac{35}{32768} \\
0 & 0 & 0 & 0 & 0 & \frac{21}{65536}
\end{pmatrix}
\qquad (19)
$$

One can recognize the first few rows of the matrix $P_{m,n}$. For the first three rows

$$
\begin{aligned}
P_{1,n} &= \frac{(2n-2)!}{(n-1)!\,n!} \frac{1}{2^{2n}} \\
P_{2,n} &= \frac{(2n-2)!}{(n-1)!\,n!} \frac{1}{2^{2n+1}} \qquad n \geq 2 \\
P_{3,n} &= \frac{(2n-4)!}{(n-3)!\,n!} \frac{3}{2^{2n+1}} \qquad n \geq 3
\end{aligned}
\qquad (20)
$$

These results are straightforwardly verified by induction. One can also identify the diagonal elements

$$P_{n,n} = \frac{(2n-2)!}{(n-1)!\,n!} \frac{1}{2^{3n-1}} \qquad (21)$$

We haven't succeeded, however, in finding an explicit expression for $P_{m,n}$ in the general case.

Consider now the three-type CBP. We again limit ourselves with the simplest three-type CBP which is represented by the scheme
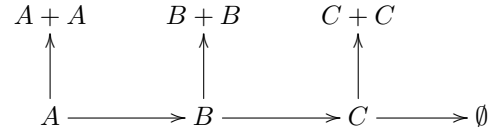
where all events occur with probabilities $\frac{1}{2}$. The probability $P_{m,n,\ell}$ that exactly (m, n, ℓ) individuals of types (A, B, C) are born before the outbreak is over is given by an obvious generalization of Eq. (13):

$$P_{m,n,\ell} = A_m \sum_{i_1 + \ldots + i_m = n} A_{i_1} \ldots A_{i_m} \sum_{j_1 + \ldots + j_n = \ell} A_{j_1} \ldots A_{j_n}$$

Converting this recurrence into an equation for the generating function

$$\mathcal{P}(x,y,z) = \sum_{\ell \geq n \geq m \geq 1} P_{m,n,\ell} x^m y^n z^\ell$$

we find

$$
\begin{aligned}
\mathcal{P}(x,y,z) &= \mathcal{A}(x\mathcal{A}(y\mathcal{A}(z))) \\
&= 1 - \sqrt{1 - x + x\sqrt{1 - y + y\sqrt{1-z}}}
\end{aligned}
$$

Generally for the $p-$type branching process the generating function

$$\mathcal{P}(z_1, \ldots, z_p) = \sum_{m_p \geq \ldots \geq m_1 \geq 1} P_{m_1, \ldots, m_p} z_1^{m_1} \ldots z_p^{m_p}$$

which encodes the joint distribution $P_{m_1,\ldots,m_p}$ can be expressed through the generating function $\mathcal{A}$ of the single type process as

$$\mathcal{P}(z_1,\ldots,z_p) = \mathcal{A}(z_1\mathcal{A}(z_2\mathcal{A}(\ldots\mathcal{A}(z_{p-1}\mathcal{A}(z_p))\ldots))) \tag{22}$$

Specializing $\mathcal{P}(z_1,\ldots,z_p)$ to $z_1 = \ldots = z_{p-1} = 1, z_p = z$ we must reproduce the generating function $\mathcal{A}^{(p)}(z)$. We actually get an identity

$$\mathcal{A}^{(p)}(z) = \mathcal{A}(\mathcal{A}(\ldots\mathcal{A}(\mathcal{A}(z))\ldots)) \tag{23}$$

Using (23) in conjunction with (3) we indeed recover (11).

III. MORE GENERAL PROCESSES

Let us first generalize the results of the previous section to non-critical brith-death processes. For a birth-death branching process we set the recovery rate to unity and denote by α the infection rate. The process is called subcritical for $\alpha < 1$, critical for $\alpha = 1$, and supercritical for $\alpha > 1$. While for $\alpha \leq 1$ the outbreak is finite with probability one; for a supercritical process ($\alpha > 1$) the outbreak never stops with probability $1 - 1/\alpha$. We can still calculate the outbreak size distribution by conditioning on finite outbreaks. Let A_n be the probability that a single individual results in a finite outbreak where exactly n individuals are infected during the epidemic. We can write the following recursion

$$A_n = \frac{\alpha}{1+\alpha} \sum_{i+j=n} A_i A_j + \frac{1}{1+\alpha} \delta_{n,1} \tag{24}$$

which, for the generating function takes the form

$$(1+\alpha)\mathcal{A} = \alpha\mathcal{A}^2 + z \tag{25}$$

and its solution is given by

$$\mathcal{A} = \frac{1+\alpha - \sqrt{(1+\alpha)^2 - 4\alpha z}}{2\alpha} \tag{26}$$

The probability of a finite outbreak is of course

$$\mathcal{A}(z=1) = \begin{cases} 1 & \text{for } \alpha \leq 1 \\ 1/\alpha & \text{for } \alpha \geq 1 \end{cases} \tag{27}$$

and the average outbreak size in case of a finite outbreak is

$$\langle n \rangle = \sum n A_n = \mathcal{A}'(z=1) = \frac{1}{|1-\alpha|} \tag{28}$$

By expanding $\mathcal{A}$ in powers of z, we obtain the probability of a finite outbreak of size n

$$A_n = \frac{1+\alpha}{4\alpha\sqrt{\pi}} \left[\frac{4\alpha}{(1+\alpha)^2} \right]^n \frac{\Gamma\left(n-\frac{1}{2}\right)}{\Gamma(n+1)} \tag{29}$$

Asymptotically, for large n we have a power law decay

$$A_n \sim \frac{1+\alpha}{4\alpha\sqrt{\pi}} e^{-n/\xi} n^{-3/2} \tag{30}$$

with an exponential cutoff around

$$\xi = [2\log(1+\alpha) - \log 4\alpha]^{-1} \tag{31}$$

The cutoff diverges as $\alpha \to 1$, and for the critical process $\alpha = 1$ we find a pure power law decay (4).

More generally, consider a general branching process where each individual is replaced by k individuals at rate α_k. Equivalently, we can consider a discrete time branching process, where at every time step each individual is independently replaced by k individuals with probability $p_k = \alpha_k / \sum \alpha_n$. In this case the governing equation for the outbreak distribution (which is essentially the backward Kolmogorov equation) is a simple generalization of Eq. (25), namely

$$\mathcal{A} \sum_{k\geq 0} \alpha_k = \sum_{k\geq 1} \alpha_k \mathcal{A}^k + \alpha_0 z \tag{32}$$

An explicit solution becomes less accessible for larger number of possible offspring, but for example the critical process with three offspring ($\alpha_0 = 2/3$, $\alpha_3 = 1/3$; other rates are zero) with a single initial individual is still tractable

$$\mathcal{A} = 2 \sin \frac{\arcsin z}{3} \tag{33}$$

Note that in this case it is impossible to find a cluster with even number of individuals: $A_{2n} \equiv 0$. The first few nonzero values of A_n are

$$A_1 = \frac{2}{3}, A_3 = \frac{8}{81}, A_5 = \frac{32}{729}, A_7 = \frac{512}{19683}, A_9 = \frac{28160}{1594323}$$

From the behavior of $\mathcal{A}(z)$ around $z = 1$ we find that for large odd n the asymptotic is

$$A_n \sim \frac{1}{\sqrt{6\pi}} n^{-3/2} \tag{34}$$

The exponent is again equal to $3/2$. Using (32) one can show that this exponent is universal for any (single-type) branching process. We state this for the equivalent discrete time version where the probability of a individual having k offspring is $p_k = \alpha_k / \sum_n \alpha_n$ for $k = 0, 1, 2, \ldots$:

Consider a discrete time branching process with a single initial individual and an offspring probability function p_k. Suppose that the process is critical, $\sum k p_k = 1$, and has finite variance $\sigma^2 = \sum k^2 p_k - 1 < \infty$. Then asymptotically $1 - \mathcal{A} \sim \sqrt{c(1-z)}$ for $z \to 1$, and

$$A_n \sim \sqrt{\frac{c}{4\pi}} n^{-3/2} \tag{35}$$

for $n \to \infty$, with $c = 2p_0/\sigma^2$.

To prove this assertion we recall that $\mathcal{A}(z)$ is analytic for $|z| < 1$ and $\mathcal{A}(1) = 1$. We write $\mathcal{A}(z) = 1 - \mathcal{B}(z)$, where $\lim_{z \to 1} \mathcal{B}(z) = 0$, and plug this expression into (32). Using the conditions for the zeroth and first moments of p_k, we find

$$
\begin{aligned}
1 - \mathcal{B} &= \sum_{k \geq 0} p_k (1 - \mathcal{B})^k - p_0 (1 - z) \\
&= 1 - \mathcal{B} + \frac{\mathcal{B}^2 \sigma^2}{2} + O(\mathcal{B}^3) - p_0(1 - z)
\end{aligned}
$$

which gives the leading asymptotic behavior

$$
\mathcal{B} = \sqrt{c(1 - z)} + O(1 - z) \tag{36}
$$

with $c = 2 p_0 / \sigma^2$. Expanding $\sqrt{1 - z}$, or equivalently using a Tauberian theorem [1], we arrive at the asymptotic (35) for A_n, which completes the proof.

The asymptotic behavior of multi-type processes now follows from recursion (23):

$$
\begin{aligned}
1 - \mathcal{A}^{(p)}(z) &= 1 - \mathcal{A}(\mathcal{A}^{(p-1)}(z)) \sim c^{1/2}[1 - \mathcal{A}^{(p-1)}(z)]^{1/2} \\
&\sim c^{1/2} c^{1/4} [1 - \mathcal{A}^{(p-2)}(z)]^{1/4} \sim \cdots \\
&\sim c^{1 - 2^{-p}} (1 - z)^{2^{-p}}
\end{aligned}
$$

If at all stages there are different critical branching processes, then simply

$$
1 - \mathcal{A}^{(p)}(z) \sim c_1^{1/2} c_2^{1/4} \cdots c_p^{1/2^p} (1 - z)^{2^{-p}} \tag{37}
$$

where all c_j are defined analogously. This corresponds to the asymptotic behavior for the probabilities

$$
A_n \propto n^{-1 - 2^{-p}} \tag{38}
$$

IV. FINITE SIZE EFFECTS

Epidemic outbreaks obviously involve finite populations. In other applications of branching processes the finiteness also plays a crucial role. Finite size effects are especially pronounced for the critical branching processes [14–16]. Therefore it is important to understand how the basic characteristics such as the average outbreak size [15, 16] vary with population size N.

The average size $\langle n \rangle = \sum_{n \geq 1} n A_n$ of the epidemic outbreak diverges in an infinite system due to the power-law tail (4) of the probability distribution A_n. Setting the upper bound at the population size N one estimates the average size of the epidemic outbreak

$$
\langle n \rangle_{\text{naive}} \propto \sum_{n=1}^{N} n^{-1/2} \propto \sqrt{N} \tag{39}
$$

This heuristic argument is actually based on the tacit assumption that a finite fraction of the population [as reflected by the upper limit in the sum in (39)] may get infected. It turns out that the actual average outbreak size is much smaller because the epidemic outbreak weakens as more individuals become infected. Let M be the maximal outbreak size. The same estimate as in (39) now gives $\langle n \rangle \propto \sqrt{M}$. On the other side, the effective infection rate is equal to $(N - M)/N = 1 - M/N$, and the average size of an outbreak in such sub-critical branching process is N/M, see (28). These two estimates should match, $\sqrt{M} \propto N/M$, from which $M \propto N^{2/3}$ thereby amending the naive estimate (39) to a self-consistent estimate [16]

$$
\langle n \rangle \propto \sqrt[3]{N} \tag{40}
$$

The scaling law (40) gives the correct asymptotic behavior. This interesting result was established by Martin-Löf [15]; several more recent studies [16–20] confirm (40) and analyze other finite-size effects.

Consider now the two-type CBP in the context of epidemic processes. We can think about As as individuals carrying light infection which can be transmitted to susceptible individuals and evolve into deadly infection. The latter can be also be transmitted to susceptible individuals or it can cause death. We need to find the scaling behavior of the maximal outbreak sizes M_A and M_B. Having established M_A and M_B, one can use the power-law tails (4) and (8) to deduce the average outbreak sizes

$$
\langle n_A \rangle \propto \sqrt{M_A}, \quad \langle n_B \rangle \propto (M_B)^{3/4} \tag{41}
$$

We again obtain an estimate $\sqrt{M_A} \propto N/M_B$. A naive conjecture is that the maximal outbreak sizes are comparable. In this situation

$$
\begin{aligned}
M_A \propto M_B &\propto N^{2/3} \\
\langle n_A \rangle \propto \sqrt[3]{N}, \quad &\langle n_B \rangle \propto \sqrt{N}
\end{aligned} \tag{42}
$$

We know, however, that different types follow different scaling laws, specifically $A_n \propto n^{-3/2}$ and $B_n \propto n^{-5/4}$, and hence it seems much more plausible that $M_A \ll M_B$. In this case we use

$$
\sqrt{M_A} \propto \frac{N}{M_B}, \quad \sqrt{M_A} \frac{N}{M_B} \propto (M_B)^{3/4} \tag{43}
$$

The second relation in (43) is understood by noting that Eq. (28) with $\alpha = 1 - M_B/N$ gives N/M_B for the average size of the outbreak caused by a single B, so in the case when the number of the seed individuals of type B is equal to $\sqrt{M_A}$, we get $\sqrt{M_A} N/M_B$ for the average size of the outbreak. From (43) we obtain the scaling of the maximal outbreak sizes

$$
M_A \propto N^{\frac{6}{11}}, \quad M_B \propto N^{\frac{8}{11}} \tag{44}
$$

which is combined with (41) to yield the average sizes

$$
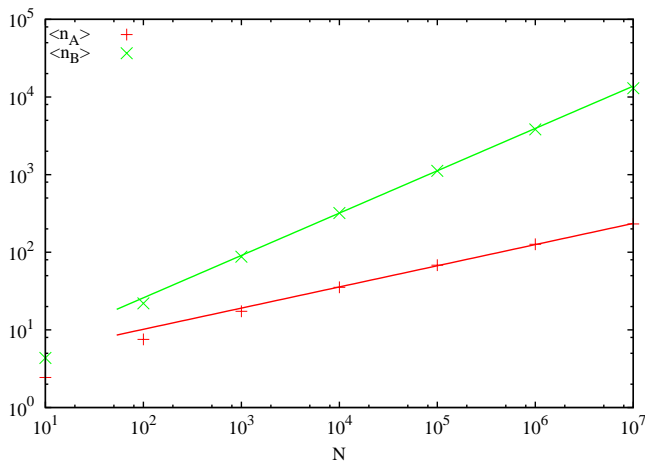\langle n_A \rangle \propto N^{\frac{3}{11}}, \quad \langle n_B \rangle \propto N^{\frac{6}{11}} \tag{45}
$$

FIG. 1: Mean outbreak sizes arising in simulations for a two-type CBP in finite populations from an average over 10^7 runs. The slopes of the lines are given by (45).
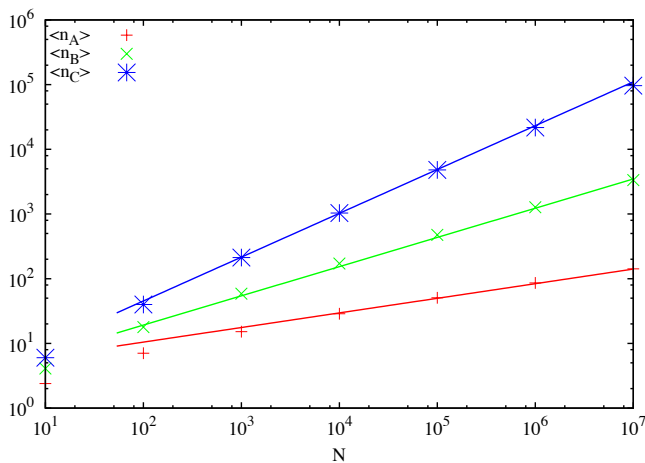

FIG. 2: Mean outbreak sizes arising in simulations for a three-type CBP in finite populations from an average over 10^6 runs. The slopes of the lines are given by (47).

For the three-type CBP the analog of (43) reads

$$\frac{N}{M_C} \propto (M_A)^{1/2}$$

$$(M_A)^{1/2}\,\frac{N}{M_C} \propto (M_B)^{3/4}$$

$$(M_B)^{3/4}\,\frac{N}{M_C} \propto (M_C)^{7/8}$$

from which the maximal outbreak sizes scale as

$$M_A \propto N^{\frac{14}{31}}, \quad M_B \propto N^{\frac{56}{63}}, \quad M_C \propto N^{\frac{24}{31}} \qquad (46)$$

and the average outbreak sizes are given by

$$\langle n_A \rangle \propto N^{\frac{7}{31}}, \quad \langle n_B \rangle \propto N^{\frac{14}{31}}, \quad \langle n_C \rangle \propto N^{\frac{21}{31}} \qquad (47)$$

The predictions (45) and (47) for the average outbreak

sizes in the two- and three-type CBPs are in excellent agreement with simulations, see Figs. 1 and 2.

Generally for the $p-$type CBP one finds that the maximal and average outbreak sizes scale as

$$M_q \propto N^{\lambda_p q/(1-2^{-q})}, \quad \langle n_q \rangle \propto N^{\lambda_p q} \qquad (48)$$

where $q = 1, 2, \ldots, p$ labels the types of individuals and

$$\lambda_p = \frac{p^{-1}(1-2^{-p})}{1+p^{-1}(1-2^{-p})}$$

The outbreaks of individuals of type p (those which die rather than differentiate) are the largest:

$$M_p \propto N^{1/[1+p^{-1}(1-2^{-p})]}$$

$$\langle n_p \rangle \propto N^{(1-2^{-p})/[1+p^{-1}(1-2^{-p})]}$$

V. DISCUSSION

We investigated the multistage SIR infection process at the epidemic threshold. In an infinitely large population, we computed the outbreak size distributions and showed that these distributions have universal algebraic tails. Combining these infinite-population results and heuristic arguments we established scaling laws for finite populations. Specifically, we showed that the maximal and average outbreak sizes scale as powers of the population size. For instance, for the three-type critical branching process six non-trivial exponents, see Eqs. (46)–(47), describe the maximal and average outbreak sizes of the three types. We numerically verified the scaling laws and found an excellent agreement.

A number of extensions are worth pursuing, especially those which would increase our understanding of the finite size scaling. Many of these extensions are easy to formulate, yet very challenging to analyze. For instance, one would like to prove the validity of the chief scaling results (48). Even more demanding would be to determine the probabilities $A_n(N), B_n(N)$, etc. For the two-type CBP, for instance, the ratios of the probabilities $A_n(N), B_n(N)$ to their infinite-population values $A_n(\infty), B_n(\infty)$ given by Eqs. (4), (8), are expected to exhibit scaling behaviors

$$\frac{A_n(N)}{A_n(\infty)} = F\!\left(\frac{n}{N^{6/11}}\right), \quad \frac{B_n(N)}{B_n(\infty)} = G\!\left(\frac{n}{N^{8/11}}\right) \quad (49)$$

The scaled sizes in Eq. (49) are fixed by the scaling laws (44) for the maximal outbreak sizes.

It would be also interesting to determine the duration, both average and maximal, of the outbreaks; for the classical critical SIR process, this problem has been investigated in [14, 16, 20].

Finally, we stress that in this paper we limited ourselves to stochastic processes without any spatial structure. Needless to say, in applications the spatial or network structure of the infected domain play an important role [21–25]. The generalizations to such settings form an interesting avenue for future research.

Appendix A: Derivation of two-type probabilities

We have obtained the explicit expression (14) by differentiating the generating function

$$P_{m,n} = \frac{1}{m!n!}\partial_x^m \partial_y^n \mathcal{P}|_{x=y=0} \qquad (A1)$$

First we differentiate with respect to x to obtain

$$\partial_x^m \mathcal{P} = \frac{(2m-3)!!}{2^m}\frac{(1-\sqrt{1-y})^m}{[1-x(1-\sqrt{1-y})]^{m-1/2}} \qquad (A2)$$

Before we differentiate with respect to y, we realize that we can take $x = 0$ already at this point, since it won't affect the result. Hence we should differentiate with respect to y only

$$\partial_x^m \mathcal{P}|_{x=0} = \frac{(2m-3)!!}{2^m}(1-\sqrt{1-y})^m \qquad (A3)$$

At each time we differentiate $(1-\sqrt{1-y})^m$ with respect to y, the number of terms grow by one until $n = m$, at which point the m terms are proportional to $(1-\sqrt{1-y})^l$ with $l = 0, \ldots, m-1$. Differentiating further won't increase the number of terms. We can write the n-th derivative as a sum

$$\partial_y^n (1-\sqrt{1-y})^m = $$
$$\sum_{k=\max(1,n-m+1)}^{n} \frac{s_{n,k}\, m!\, (1-\sqrt{1-y})^{m-n-1+k}}{2^n(m-n-1+k)!\,(1-y)^{(n+k-1)/2}}$$

This can be easily proved by induction and by using the recursion formula (17) for $s_{n,k}$. Setting $y = 0$ in the sum on the right-hand side, we are left with only one term (the smallest k term). Accounting for all factors in (A1) we arrive at the announced result (16).

[1] W. Feller, *An Introduction to Probability Theory and Its Applications*, Vol. I, 3rd edn. (John Wiley, New York, 1968).

[2] T. E. Harris, *The Theory of Branching Processes* (Dover, New York, 1989).

[3] K. B. Athreya and P. E. Ney, *Branching Processes* (Dover Publications, Inc., Mineola, New York, 2004).

[4] M. Kimmel and D. Axelrod, *Branching Processes in Biology* (Springer, New York, 2002).

[5] P. Haccou, P. Jagers, and V. A. Vatutin, *Branching Processes: Variation, Growth, and Extinction of Populations* (Cambridge University Press, New York, 2005).

[6] N. T. J. Bailey, *The Mathematical Theory of Epidemics* (New York, Hafner, 1957); an extended and updated version of this book is N. T. J. Bailey, *The Mathematical Theory of Infectious Diseases* (Oxford University Press, Oxford, 1987).

[7] R. Anderson and R. May, *Infectious Diseases: Dynamics and Control* (Oxford University Press, Oxford, 1991).

[8] H. W. Hethcote, SIAM Rev. **42**, 599 (2000).

[9] J. D. Murray, *Mathematical Biology. I. An Introduction* (Spriger-Verlag, New York, 2002).

[10] L. Popovic, Ann. Appl. Probab. **14**, 2120–2148 (2004); D. Aldous and L. Popovic, Adv. Appl. Probab. **37**, 1094–1115 (2005); T. Gernhard, Bull. Math. Biol. **70**, 1082–1097 (2008); D. J. Aldous, M. A. Krikun, and L. Popovic, Syst. Biol. **60**, 318–328 (2011).

[11] S. Nee, Paleobiology **30**, 172–178 (2004); Annu. Rev. Ecol. Evol. Syst. **37**, 1–17 (2006).

[12] R. Antia, R. R. Regoes, J. C. Koella, and C. T. Bergstrom, Nature **426**, 658 (2003).

[13] T. Antal and P. L. Krapivsky, J. Stat. Mech. P07028 (2010); J. Stat. Mech. P08018 (2011).

[14] C. J. Ridler-Rowe, J. Appl. Probab. **4**, 19 (1967).

[15] A. Martin-Löf, J. Appl. Probab. **35**, 671 (1998).

[16] E. Ben-Naim and P. L. Krapivsky, Phys. Rev. E **69**, 050901(R) (2004).

[17] D. A. Kessler and N. M. Shnerb, Phys. Rev. E **76**, 010901(R) (2007).

[18] L. F. Gordillo, S. A. Marion, A. Martin-Löf, and P. E. Greenwood, Bull. Math. Biol. **70**, 589 (2008).

[19] R. van der Hofstad, A. J. E. M. Janssen, and J. S. H. van Leeuwaarden, Adv. Appl. Probab. **42**, 1187 (2010).

[20] E. Ben-Naim and P. L. Krapivsky, arXiv:1202.1853.

[21] P. Grassberger, Math. Biosci. **63**, 157 (1983).

[22] J.-F. Le Gall, *Spatial Branching Processes, Random Snakes and Partial Differential Equations* (Birkhäuser, Basel, 1999).

[23] J. D. Murray, *Mathematical Biology. II. Spatial Models and Biomedical Applications* (Springer-Verlag, New York, 2002).

[24] G. M. Schütz, M. Brandaut, and S. Trimper, Phys. Rev. E **78**, 061132 (2008).

[25] T. Tomé and R. M. Ziff, Phys. Rev. E **82**, 051921 (2010).