

Approximate modeling for supercritical Galton-Watson branching processes with compound Poisson-gamma distribution

Kyoya Uemura ^{*}, Tomoyuki Obuchi , and Toshiyuki Tanaka 

Graduate School of Informatics, [Kyoto University](https://www.kyoto-u.ac.jp/), Kyoto 606-8501, Japan



(Received 26 October 2025; accepted 13 January 2026; published 2 March 2026)

We study asymptotic properties of supercritical Galton-Watson (GW) branching processes in the asymptotic where the mean of the offspring distribution approaches 1 from above. We show that the population-size distribution of the GW branching processes at a sufficiently large generation in this asymptotic can be approximated by a compound Poisson-gamma distribution. Numerical experiments revealed that the compound Poisson-gamma models were in good agreement with the corresponding GW models for sufficiently large generations under a reasonable parameter regime. Our results can be regarded as supporting the use of the compound Poisson-gamma model as a model for cascaded multiplication processes.

DOI: [10.1103/w3yn-jpt5](https://doi.org/10.1103/w3yn-jpt5)

I. INTRODUCTION

A Galton-Watson (GW) branching process (also called a Bienaymé-Galton-Watson branching process) is a stochastic model that describes population evolution over generations, which was originally developed to study propagation and extinction probabilities of family names [1,2]. In the GW process, each individual in the n th generation independently produces a random number of offspring in the $(n+1)$ th generation. Let Z_n denote the population size of n th generation. Then, $(Z_n)_{n=0,1,\dots}$ evolves as follows:

$$Z_{n+1} = \sum_{i=1}^{Z_n} X_n^{(i)}, \quad n = 0, 1, \dots, \quad (1)$$

where $X_n^{(i)}$ denotes the number of offspring produced by the individual $i \in \{1, \dots, Z_n\}$ in the n th generation, and where $\{X_n^{(i)}\}$ are independent and identically distributed (i.i.d.) according to a certain offspring distribution with mean λ . Unless otherwise stated, we assume $Z_0 \equiv 1$. When $\lambda > 1$ (referred to as the supercritical case), the GW process provides a reasonable model for cascaded multiplication processes such as electron multiplication, where the population increases with expected gain λ in each generation. As a result, the supercritical GW process is now widely applied to model the population dynamics of these multiplication processes in fields such as mathematical biology and physics [3–12].

One of the main interests in the supercritical GW process is regarding the distribution of Z_n when n is large. For

instance, in the context of electron multiplication, this distribution can be interpreted as the single electron response (SER) distribution (also known as the pulse height distribution) of an electron multiplier (EM) with a sufficient number of dynodes, providing insights into properties of response signals of detectors using EMs. A primary approach for studying the distribution of Z_n for large n is to consider an alternative random variable defined by normalizing Z_n with the total gain $\mathbb{E}[Z_n] = \lambda^n$ up to the n th generation [13,14]. Let

$$W_n = \frac{Z_n}{\lambda^n}. \quad (2)$$

One then has

$$\begin{aligned} \mathbb{E}[W_{n+1} \mid W_1, \dots, W_n] &= \lambda^{-n-1} \mathbb{E}[Z_{n+1} \mid Z_1, \dots, Z_n] \\ &= \lambda^{-n-1} \mathbb{E}[Z_{n+1} \mid Z_n] \\ &= \lambda^{-n} Z_n = W_n, \end{aligned} \quad (3)$$

which shows that the sequence $(W_n)_{n=0,1,\dots}$ is a non-negative martingale, and consequently, the martingale convergence theorem [15] tells us that, as $n \rightarrow \infty$, $(W_n)_{n=0,1,\dots}$ converges to a random variable W almost surely. In the supercritical case, the limiting distribution $\mathbb{P}(W)$ is nondegenerate under certain mild conditions [14,16] (see Sec. II A), providing an indirect evaluation of the distribution of Z_n for large n .

However, the limiting distribution $\mathbb{P}(W)$ is generally not available in a closed form, even when the offspring distribution is simple, such as the Poisson offspring distribution. This difficulty makes it challenging to apply the GW process model to downstream data analyses. In practice, more tractable models based on empirical insights into the dataset (e.g., non-negativity, overdispersion, and excess zeros) are commonly adopted [17–19]. For example, Bellamy *et al.* [17] proposed describing the operation of a photomultiplier as consisting of two independent processes: photodetection and amplification. They then formulated the response of a photomultiplier by using a compound Poisson model, where

^{*}Contact author: uemura.kyoya.n08@kyoto-u.jp

the number of photoelectrons on the photodetector is modeled by a Poisson distribution, and the amplification process is characterized by an SER distribution. Subsequent studies [20,21] discussed heuristic modeling of the SER distribution using a gamma distribution, resulting in a compound Poisson-gamma (CPG) distribution as a model for the photomultiplier response. However, the theoretical basis of these modeling approaches has rarely been discussed.

In this work, we study the asymptotics of supercritical GW processes as they approach criticality, that is, as the mean λ of the offspring distribution approaches 1 from above (i.e., $\lambda \downarrow 1$). In the asymptotic regime $\lambda \downarrow 1$ with $Z_0 \equiv 1$ (Condition I), we derive a probability distribution in closed form whose cumulant generating function matches that of the limiting distribution $\mathbb{P}(W)$ up to the leading order in $(\lambda - 1)$. It is also shown that the derived distribution belongs to a class of CPG distributions. This is significant from a practical perspective, as the CPG distributions can be reparametrized as the Tweedie family, a subclass of exponential dispersion models [22], which are known to possess several practically desirable properties, similar to those of exponential families [23,24]. Moreover, we present a similar result for the case where Z_0 follows a certain distribution with mean λ_0 , where λ_0 does not need to be close to unity (Condition II). Note that Condition II reflects the standard assumption in response signal modeling of EMs [9,10], in which a particle incident to the first-stage dynode produces secondary electrons with a gain different from that of the subsequent stages. Numerical experiments revealed that these CPG models were in good agreement with the corresponding GW models for large n under a reasonable parameter regime. These results provide mathematical and empirical support for the use of CPG distributions as a model for cascaded multiplication processes.

The remainder of this paper is organized as follows. Section II discusses our main theoretical results, which states that the limiting distribution $\mathbb{P}(W)$ in the asymptotic $\lambda \downarrow 1$ can be approximated by a CPG distribution under both Conditions I and II. Section III verifies the main results via comparing the population-size distribution $\mathbb{P}(Z_n)$ with large n and the CPG distribution. Section IV discusses tail behaviors of the limiting distribution $\mathbb{P}(W)$. It also discusses the diffusion approximation and a large- n analysis, both of which further support the use of the CPG distribution as a model of the population-size distribution of the GW processes when n is large. Section V concludes this paper.

II. THEORETICAL RESULTS

A. Condition I: $Z_0 \equiv 1$

In this subsection, we discuss the limiting distribution for the supercritical GW process when $Z_0 \equiv 1$ (Condition I). Let $f(s)$ and $\phi(t)$ be the probability generating function and the moment generating function of the offspring distribution, respectively,

$$f(s) = \mathbb{E}[s^{X_n^{(i)}}] = \sum_{k=0}^{\infty} \mathbb{P}(X_n^{(i)} = k) s^k, \quad (4)$$

$$\phi(t) = \mathbb{E}[e^{tX_n^{(i)}}] = f(e^t). \quad (5)$$

Let $f_n(s)$ denote the probability generating function of Z_n . From Eq. (1), one has

$$\begin{aligned} f_{n+1}(s) &= \mathbb{E}_{Z_{n+1}}[s^{Z_{n+1}}] \\ &= \mathbb{E}_{Z_n}[\{\mathbb{E}_{X_n^{(i)}}[s^{X_n^{(i)}}]\}^{Z_n}] (\because \{X_n^{(i)}\} \text{ is i.i.d.}) \\ &= f_n(f(s)), \quad n \geq 1, \end{aligned} \quad (6)$$

where

$$f_0(s) = s, \quad f_1(s) = f(s). \quad (7)$$

By recursively applying Eq. (6), one can in principle obtain the probability generating function $f_n(s)$ of Z_n for any n . In practice, however, for large n its formula has a higher-order nested structure as

$$f_n(s) = f^n(s) = \overbrace{(f \circ f \circ \cdots \circ f)}^{n \text{ times}}(s), \quad (8)$$

and this mathematical complexity hinders the application of the GW model to general data analysis unless f has a very specific structure [13,25,26].

Here, to evaluate the probability distribution of Z_n for large n , we follow the classical approach in the theory of GW branching processes and begin our discussion by considering the normalized random variable W_n defined by Eq. (2), instead of Z_n . According to the Kesten-Stigum theorem [16] (see also Ref. [14], Theorem I.10.1), when $\lambda > 1$, the sequence $(W_n)_{n=0,1,\dots}$ converges almost surely (a.s.) to a nondegenerate random variable W , if and only if $\mathbb{E}[X_n^{(i)} \ln X_n^{(i)}] < \infty$ holds. For the limiting random variable W , one has $\mathbb{E}[W] = 1$, and the moment generating function $M_W(t)$ of W satisfies the following functional equation:

$$M_W(\lambda t) = f(M_W(t)). \quad (9)$$

It can further be rewritten in terms of the cumulant generating function $K_W(t) = \ln M_W(t)$ of W as

$$K_W(\lambda t) = \ln f(e^{K_W(t)}) = \psi(K_W(t)), \quad (10)$$

where $\psi(t) = \ln \phi(t) = \ln f(e^t)$ denotes the cumulant generating function of the offspring distribution. These functional equations lay the basis for studying properties of the limiting random variable W .

One can notice that the functional equation (10) is invariant under scaling of t , which implies that the cumulant generating function of aW for an arbitrary $a > 0$ satisfies the same functional equation. With a modest amount of foresight, we consider the rescaled random variable $\bar{W}$ defined as

$$\bar{W} = (\lambda - 1)W, \quad (11)$$

for which $\mathbb{E}(\bar{W}) = \lambda - 1$, and study properties of $\bar{W}$ via the following functional equation corresponding to Eq. (10):

$$\bar{K}(\lambda t) = \psi(\bar{K}(t)), \quad (12)$$

where $\bar{K}(t)$ denotes the cumulant generating function of $\bar{W}$.

Solving Eq.(12) analytically is intractable for most choices of the offspring distribution. We therefore resort to the perturbation method [27,28] to study the solution of Eq. (12) as $\lambda \downarrow 1$. Let

$$\epsilon = \lambda - 1 \quad (13)$$

be the perturbation parameter, and assume that, for a sufficiently small $\epsilon \geq 0$, the cumulant generating function $\psi(t)$ of the offspring distribution is expanded in ϵ as

$$\psi(t) = \psi_0(t) + \epsilon \psi_1(t) + \mathcal{O}(\epsilon^2), \quad (14)$$

where

$$\psi_0(t) = \sum_{k=1}^{\infty} \frac{\kappa_k^*}{k!} t^k = t + \sum_{k=2}^{\infty} \frac{\kappa_k^*}{k!} t^k \quad (15)$$

is the cumulant generating function of the offspring distribution at $\epsilon = 0$ (i.e., at $\lambda = 1$), where $\kappa_1^* = \mathbb{E}[X_n^{(i)}] = \lambda = 1$ holds. We also assume that κ_k^* for $k \geq 2$ are such that the power series defining $\psi_0(t)$ is convergent in a neighborhood of $t = 0$. On the other hand, for $\epsilon \geq 0$ one should have

$$\mathbb{E}[X_n^{(i)}] = 1 + \epsilon = \psi'(0) = \psi'_0(0) + \epsilon \psi'_1(0) + \mathcal{O}(\epsilon^2), \quad (16)$$

which yields $\psi'_1(0) = 1$. Then, Eq. (14) can be rewritten as

$$\psi(t) = \psi_0(t) + \epsilon(t + \tilde{\psi}_1(t)) + \mathcal{O}(\epsilon^2), \quad (17)$$

where $\tilde{\psi}_1(t) = \mathcal{O}(t^2)$ as $t \rightarrow 0$.

It should be noted that our discussion so far assumes the cumulant generating function $\psi(t)$ of the offspring distribution is well-defined for $\epsilon \geq 0$. It then implies $\mathbb{E}[X_n^{(i)} \ln X_n^{(i)}] < \infty$ to hold, which in turn ensures existence of a nondegenerate limit W of the sequence $(W_n)_{n=0,1,\dots}$ via the Kesten-Stigum theorem.

We next consider expansion of the cumulant generating function $\bar{K}(t)$ of $\bar{W}$ in ϵ . When $\epsilon = 0$, one has $\bar{W} = \epsilon W = 0$ with probability 1, yielding $\bar{K}(t) \equiv 0$ which satisfies Eq. (12) trivially. We can therefore assume the following perturbative expansion:

$$\bar{K}(t) = \epsilon \bar{K}_1(t) + \epsilon^2 \bar{K}_2(t) + \mathcal{O}(\epsilon^3), \quad (18)$$

where $\bar{K}_1(t)$ and $\bar{K}_2(t)$ denote the first- and second-order coefficients, respectively. We note that for any $\epsilon \geq 0$ one has $\bar{K}'(0) = \mathbb{E}[\bar{W}] = \epsilon$, which implies $\bar{K}'_1(0) = 1$. Substituting Eq. (18) into the left-hand side of Eq. (12) yields, up to the second order in ϵ ,

$$\begin{aligned} \bar{K}(\lambda t) &= \bar{K}((1 + \epsilon)t) \\ &= \bar{K}(t) + \epsilon t \bar{K}'(t) + \frac{\epsilon^2 t^2}{2} \bar{K}''(t) + \mathcal{O}(\epsilon^3) \\ &= \epsilon \bar{K}_1(t) + \epsilon^2 (\bar{K}_2(t) + t \bar{K}'_1(t)) + \mathcal{O}(\epsilon^3). \end{aligned} \quad (19)$$

Using Eqs. (17) and (18), the right-hand side of Eq. (12) is calculated, up to the second order in ϵ , as

$$\begin{aligned} \psi(\bar{K}(t)) &= \psi_0(\epsilon \bar{K}_1(t) + \epsilon^2 \bar{K}_2(t) + \mathcal{O}(\epsilon^3)) \\ &\quad + \epsilon(\epsilon \bar{K}_1(t) + \mathcal{O}(\epsilon^2)) + \mathcal{O}(\epsilon^3) \\ &= \epsilon \bar{K}_1(t) + \epsilon^2 \left(\bar{K}_2(t) + \frac{\kappa_2^*}{2} \{\bar{K}_1(t)\}^2 + \bar{K}_1(t) \right) \\ &\quad + \mathcal{O}(\epsilon^3). \end{aligned} \quad (20)$$

Equating (19) and (20), the zeroth-order terms vanish on both sides, and the first-order terms coincide. Comparing the coefficients of ϵ^2 yields the following ordinary differential

equation for $\bar{K}_1(t)$:

$$t \bar{K}'_1(t) = \frac{\kappa_2^*}{2} \{\bar{K}_1(t)\}^2 + \bar{K}_1(t). \quad (21)$$

We solve it under the condition $\bar{K}'_1(0) = 1$, yielding

$$\bar{K}_1(t) = \frac{t}{1 - \frac{\kappa_2^* t}{2}}. \quad (22)$$

Therefore, the perturbative approximation to the solution of the functional equation (12) as $\epsilon \downarrow 0$ (i.e., $\lambda \downarrow 1$), up to the first order in ϵ , is obtained as

$$\bar{K}(t) \approx \epsilon \bar{K}_1(t) = \frac{(\lambda - 1)t}{1 - \frac{\kappa_2^* t}{2}} =: \tilde{K}(t). \quad (23)$$

In view of Eq. (18), $\epsilon \bar{K}_1(t)$ is the leading-order term in the expansion of $\bar{K}(t)$ in ϵ , and therefore it reproduces all the cumulants of $\bar{W}$ up to the leading order in ϵ . It should be emphasized that the same result (23) can also be derived rigorously under weaker assumptions, without resorting to the perturbation method; the details are provided in Sec. S1 of the Supplemental Material [29].

From the preceding argument, as $\lambda \downarrow 1$, the cumulant generating function of $W = \bar{W}/(\lambda - 1)$ can be approximated as

$$K_W(t) \approx \tilde{K}\left(\frac{t}{\lambda - 1}\right) = \frac{t}{1 - \frac{\kappa_2^*}{2(\lambda - 1)}t}. \quad (24)$$

A remarkable fact about the result is that the asymptotic cumulant generating function $\tilde{K}(t)$ is universal, in the sense that it depends on the offspring distribution only via its variance κ_2^* in the limit $\lambda \downarrow 1$.

We next show that the cumulant generating function $\tilde{K}(\frac{t}{\lambda - 1})$ in Eq. (24) coincides with that of the CPG distribution via an appropriate reparametrization. A random variable S following a CPG distribution is defined as

$$S = \sum_{i=1}^N Y_i, \quad (25)$$

$$N \sim \text{Poisson}(\mu); \quad Y_i \stackrel{\text{i.i.d.}}{\sim} \text{Gamma}(\alpha, \tau), \quad (26)$$

where $\text{Poisson}(\mu)$ denotes a Poisson distribution with mean μ , and where $\text{Gamma}(\alpha, \tau)$ denotes a gamma distribution with shape parameter α and scale parameter τ . Via a slight abuse of notation [30], we let $P_{\text{CPG}}(\cdot; \mu, \alpha, \tau)$ denote the probability density function of the CPG distribution with parameters (μ, α, τ) . The cumulant generating function $K_S(t)$ of S is given by

$$\begin{aligned} K_S(t) &= \ln(\mathbb{E}_S[e^{tS}]) \\ &= \ln(\mathbb{E}_N[\{\mathbb{E}_{Y_i}[e^{tY_i}]\}^N]) \\ &= \ln(\mathbb{E}_N[e^{N K_{Y_i}(t)}]) \\ &= K_N(K_{Y_i}(t)) \\ &= \mu\{(1 - \tau t)^{-\alpha} - 1\}, \end{aligned} \quad (27)$$

where $K_N(t)$ and $K_{Y_i}(t)$ are the cumulant generating functions of N and Y_i , given by

$$K_N(t) = \mu(e^t - 1), \quad (28)$$

$$K_{Y_i}(t) = \ln(1 - \tau t)^{-\alpha}, \quad (29)$$

respectively. Comparing Eq. (27) with Eq. (24), we find that the cumulant generating function (24) coincides with that of the CPG distribution $P_{\text{CPG}}(\cdot; \mu, \alpha, \tau)$ with

$$\mu = \frac{2(\lambda - 1)}{\kappa_2^*}, \quad \alpha = 1, \quad \tau = \frac{\kappa_2^*}{2(\lambda - 1)}. \quad (30)$$

Consequently, it is shown that in the asymptotic regime $\lambda \downarrow 1$ the limiting distribution $\mathbb{P}(W)$ of the GW process on Condition I can be approximated by the CPG distribution with the parameters (30).

B. Condition II: Random Z_0

Motivated by practical applications, in this subsection, we consider the case where Z_0 follows a certain distribution $\mathbb{P}(Z_0)$ with mean λ_0 (Condition II). Note that, unlike λ , λ_0 is not required to be close to unity.

Let us assume that the cumulant generating function $K_{Z_0}(t)$ of Z_0 exists:

$$K_{Z_0}(t) = \sum_{v=1}^{\infty} \frac{\kappa_{Z_0,v}}{v!} t^v, \quad (31)$$

where $\kappa_{Z_0,v}$ denotes the v th cumulant of Z_0 , and $\kappa_{Z_0,1} = \lambda_0$. Then, consider a random variable V_n defined by

$$V_n = \frac{Z_n}{\lambda^n}. \quad (32)$$

Applying the result of the discussion in Sec. II A, as $n \rightarrow \infty$ and $\lambda \downarrow 1$, one can approximately model the limiting distribution of V_n as

$$V = \sum_{i=1}^{Z_0} W^{(i)}, \quad (33)$$

$$Z_0 \sim \mathbb{P}(Z_0), \quad (34)$$

where V denotes the limit of V_n as $n \rightarrow \infty$. $\{W^{(i)}\}$ denotes a set of i.i.d. random variables that follow the CPG distribution with parameters (30), that is, for any i ,

$$W^{(i)} = \sum_{j=1}^{N^{(i)}} \Gamma_j^{(i)}, \quad (35)$$

where

$$N^{(i)} \stackrel{\text{i.i.d.}}{\sim} \text{Poisson}\left(\frac{2(\lambda - 1)}{\kappa_2^*}\right), \quad (36)$$

$$\Gamma_j^{(i)} \stackrel{\text{i.i.d.}}{\sim} \text{Gamma}\left(1, \frac{\kappa_2^*}{2(\lambda - 1)}\right). \quad (37)$$

Here, let

$$N_0 = \sum_{i=1}^{Z_0} N^{(i)}. \quad (38)$$

Then, Eq. (33) can be rewritten using N_0 as

$$V = \sum_{j=1}^{N_0} \Gamma_j, \quad (39)$$

$$\Gamma_j \stackrel{\text{i.i.d.}}{\sim} \text{Gamma}\left(1, \frac{\kappa_2^*}{2(\lambda - 1)}\right). \quad (40)$$

From Eqs. (36) and (38), the cumulant generating function $K_{N_0}(t)$ of N_0 can be obtained in a similar manner to Eq. (27) as follows:

$$\begin{aligned} K_{N_0}(t) &= K_{Z_0}(K_{N^{(i)}}(t)) \\ &= K_{Z_0}\left(\frac{2(e^t - 1)}{\kappa_2^*}(\lambda - 1)\right), \end{aligned} \quad (41)$$

where $K_{N^{(i)}}(t)$ is the cumulant generating function of the Poisson random variable $N^{(i)}$ [Eq. (36)]. Noting our assumption $\lambda - 1 \ll 1$, from the Taylor series expansion of $K_{N_0}(t)$ in the quantity $\frac{2(e^t - 1)}{\kappa_2^*}(\lambda - 1)$, one has, for any t ,

$$\lim_{\lambda \downarrow 1} \frac{K_{N_0}(t)}{\lambda - 1} = \frac{2\lambda_0}{\kappa_2^*}(e^t - 1). \quad (42)$$

Thus, as $\lambda \downarrow 1$,

$$K_{N_0}(t) = \frac{2\lambda_0(\lambda - 1)}{\kappa_2^*}(e^t - 1), \quad (43)$$

which is nothing but the cumulant generating function of the Poisson distribution with mean $\frac{2\lambda_0(\lambda - 1)}{\kappa_2^*}$. Therefore, one has

$$N_0 \sim \text{Poisson}\left(\frac{2\lambda_0(\lambda - 1)}{\kappa_2^*}\right) \quad (44)$$

in the limit $\lambda \downarrow 1$. From Eqs. (39), (40), and (44), it is shown that as $\lambda \downarrow 1$, the limiting distribution $\mathbb{P}(V)$ can be approximated by the CPG distribution $P_{\text{CPG}}(\cdot; \mu, \alpha, \tau)$ with

$$\mu = \frac{2\lambda_0(\lambda - 1)}{\kappa_2^*}, \quad \alpha = 1, \quad \tau = \frac{\kappa_2^*}{2(\lambda - 1)}. \quad (45)$$

It should be noted that the above CPG distribution approximation of $\mathbb{P}(V)$ depends on the distribution of Z_0 only via its mean λ_0 .

III. NUMERICAL VERIFICATION

A. Overview and computational issues

Here, we examined the above theoretical argument by comparing numerically evaluated distributions of Z_n for large n with the CPG distributions. At the origin, since both Z_n and the CPG random variable have probability mass, they can be compared as

$$\mathbb{P}(Z_n = 0) \approx P_{\text{CPG}}(0; \mu, \alpha, \tau) = e^{-\mu}, \quad (46)$$

where (μ, α, τ) are given by Eq. (30) under Condition I and by Eq. (45) under Condition II. On the other hand, on the interval $(0, \infty)$, a direct comparison of these distributions as above is not possible because the CPG random variable (and W) is absolutely continuous, whereas Z_n is integer valued and hence discrete. Thus, we compare them in terms of probability values via appropriate scaling as follows:

$$\mathbb{P}(Z_n = \ell) \approx \frac{1}{\lambda^n} P_{\text{CPG}}\left(\frac{\ell}{\lambda^n}; \mu, \alpha, \tau\right) \quad (47)$$

for $\ell \in \mathbb{N}$, or equivalently,

$$P_{\text{CPG}}(x; \mu, \alpha, \tau) \approx \lambda^n \mathbb{P}(Z_n = \lambda^n x), \quad (48)$$

for x such that $\lambda^n x \in \mathbb{N}$, where (μ, α, τ) are given by Eq. (30) under Condition I and by Eq. (45) under Condition II.

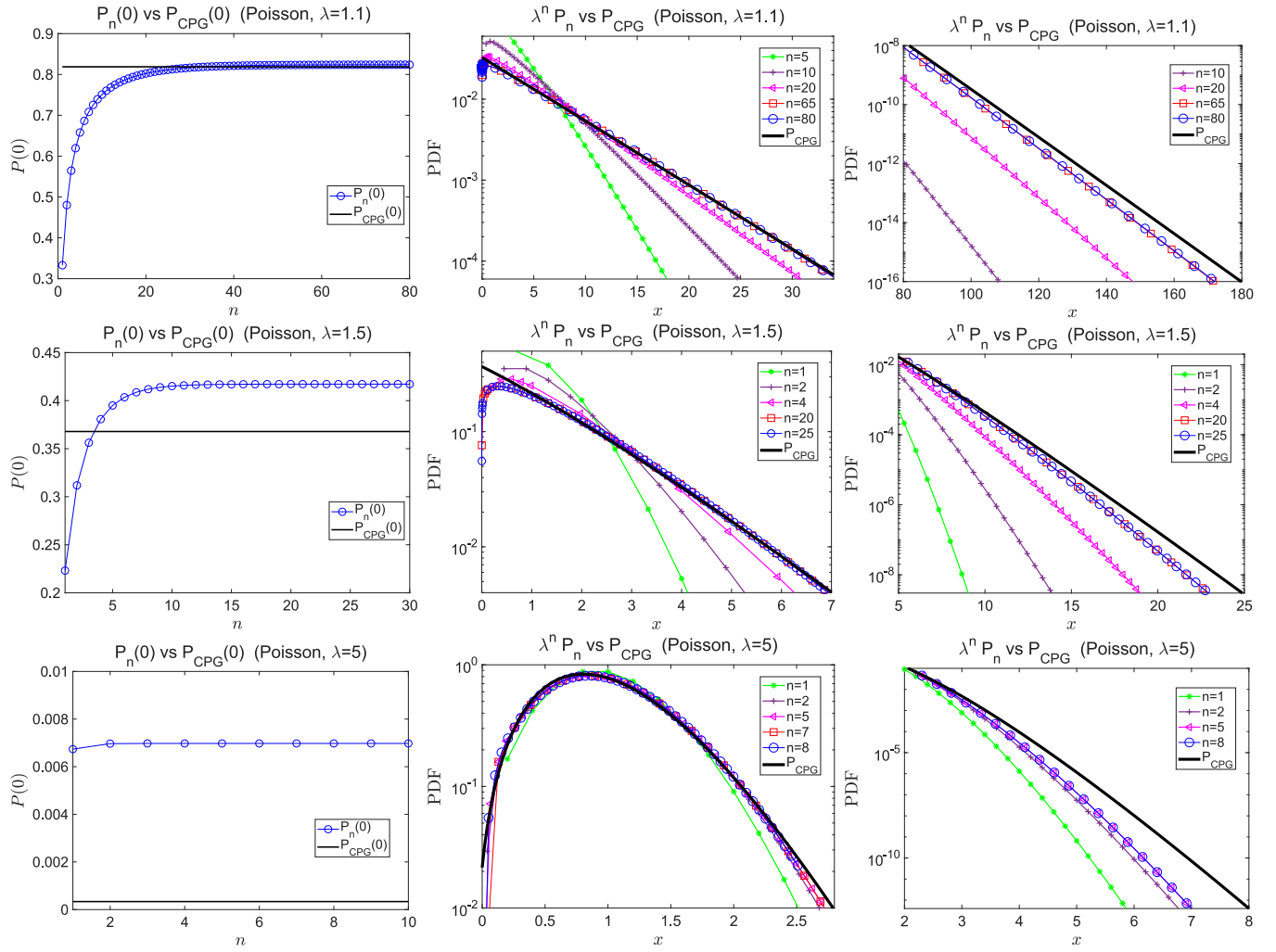


FIG. 1. Comparison between P_{CPG} and P_n in the Poisson case under Condition I. Three different values of λ , 1.1, 1.5, 5.0, are examined from top to bottom. The left column is for the plot of $P_n(0)$ (joined blue circles) against n , compared with $P_{\text{CPG}}(0)$ (black line). The other columns compare $\lambda^n P_n(\lambda^n x)$ and $P_{\text{CPG}}(x)$ in the semilogarithmic scale. The bulk region and the right tail of the distribution are exhibited in the center and right columns, respectively.

We performed quantitative verifications of these relations (46) and (48) in this section. We investigated both the Poisson and geometric offspring distributions in order to demonstrate the universality of our results with respect to the choice of the offspring distribution.

For the Poisson offspring distribution, an efficient recursion formula to compute the distribution of Z_n is known [9,10], and we employed it for computing $\mathbb{P}(Z_n = \ell)$. For the sake of self-containedness, we provided the derivation of the recursion formula for Condition I in Appendix A and for Condition II in Appendix B, where we extended it to include certain negative binomial offspring distributions. When computing $\mathbb{P}(Z_n = \ell)$ with the recursion formulas, the computational cost rapidly increases as n grows, which imposes a limitation on the practically manageable support size of the distribution. Unless otherwise specified, we considered the range $\ell \leq \ell_{\max}$ with $\ell_{\max} = 10^6$ for the Poisson case. Related to this computational issue, it should be noted that the maximum value of n for which the simulation remains sufficiently accurate heavily depends on both the support size and the value of λ . In the results below, we show

only the range of n for which $\sum_{\ell=0}^{\ell_{\max}} \mathbb{P}(Z_n = \ell) > 0.99$ is satisfied.

Meanwhile, for the geometric offspring distribution, the evaluation of $\mathbb{P}(Z_n = \ell)$ is much easier, since the probability generating function f_n takes a linear fractional form for any generation n , allowing its parameters to be computed in a simple iterative manner; the distribution of Z_n is calculated via a series expansion of the linear fractional form with the computed parameters [25]. An outline of the calculations for this case is provided in Appendix C. Thanks to this method, the computational difficulty encountered in the Poisson case is significantly alleviated in the geometric case.

B. Results under Condition I

We let $P_n(\ell)$ denote the probability $\mathbb{P}(Z_n = \ell)$ under Condition I. The Poisson and geometric offspring distributions were considered.

The results for the Poisson offspring distribution with $\lambda \in \{1.1, 1.5, 5.0\}$ are shown in Fig. 1. The top row shows results with $\lambda = 1.1$, that is, for the case where λ is close

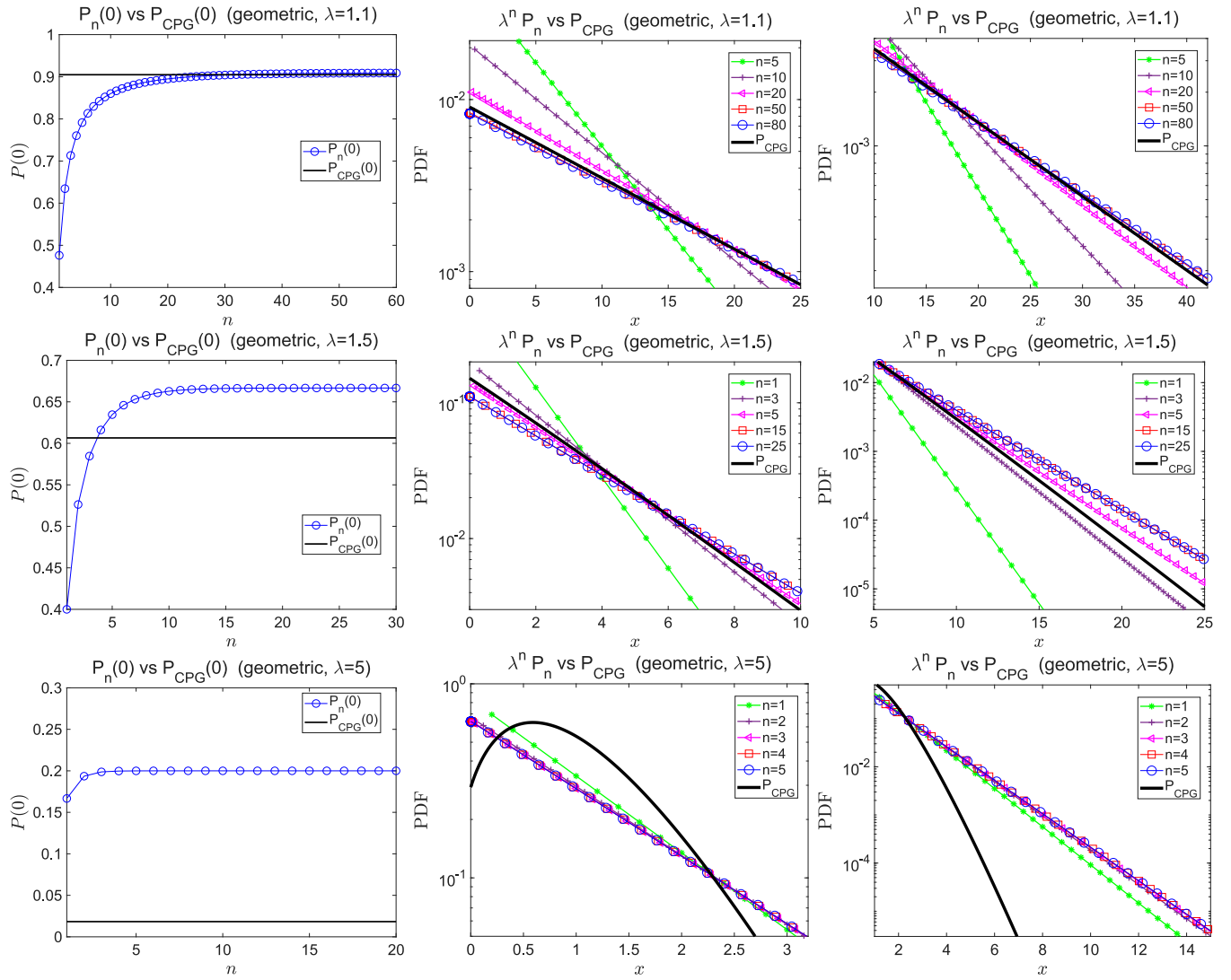


FIG. 2. Comparison between P_{CPG} and P_n in the geometric case under Condition I. Three different values of λ , 1.1, 1.5, 5.0, are examined from top to bottom. The left column is for the plot of $P_n(0)$ (joined blue circles) against n , compared with $P_{\text{CPG}}(0)$ (black line). The other columns compare $\lambda^n P_n(\lambda^n x)$ (joined color markers) and $P_{\text{CPG}}(x)$ (black line) in the semi-logarithmic scale. The bulk region and the right tail of the distribution are exhibited in the center and right columns, respectively.

to the critical value 1. One can observe that, as n grows, $P_n(0)$ closely approaches $P_{\text{CPG}}(0)$ (left panel) and that $P_{\text{CPG}}(x)$ shows good agreement with $\lambda^n P_n(\lambda^n x)$ over a fairly wide range of x for large n (center panel). These results are consistent with the theoretical argument above. Meanwhile, the right tail of $\lambda^n P_n(\lambda^n x)$ decays faster than the exponential decay of $P_{\text{CPG}}(x)$ (right panel). This is natural, as will be discussed in Sec. IV A, because the limiting random variable W of the GW process with an offspring distribution having a lighter-than-exponential tail, such as the Poisson distribution, should also have a lighter-than-exponential tail. Although the tail behaviors differ, the bulk region with non-negligible probability shows good agreement, supporting the validity of our approximation approach.

The middle and bottom rows of Fig. 1 show results for the Poisson offspring distribution with $\lambda = 1.5$ and 5.0, respectively. In both cases, the bulk regions of $P_{\text{CPG}}(x)$ still exhibit fairly good agreement with $\lambda^n P_n(\lambda^n x)$, as shown in the center

column. However, from the left column of Fig. 1, it was found that the CPG model underestimates the probability of zero of the GW model for large n . We theoretically confirmed that this underestimation problem occurs for any $\lambda > 1$ (see Sec. S2 of the Supplemental Material [31]).

The results for the geometric offspring distribution with $\lambda \in \{1.1, 1.5, 5.0\}$ are shown in Fig. 2. It is exactly the counterpart of Fig. 1 and the corresponding panel plots the same quantity at the same parameter as that in Fig. 1: only the difference is in its offspring distribution. The top row again shows good agreement between $P_{\text{CPG}}(x)$ and $\lambda^n P_n(\lambda^n x)$ at large n , demonstrating the universality of our theory when λ is close to 1. On the other hand, the tail part of the distributions exhibits an exponential decay, in contrast to the Poisson case. This is because the geometric distribution has an exponential tail: The limiting random variable W of the GW process with an offspring distribution having an exponential tail has an exponential tail as well (see Sec. IV A for a detailed

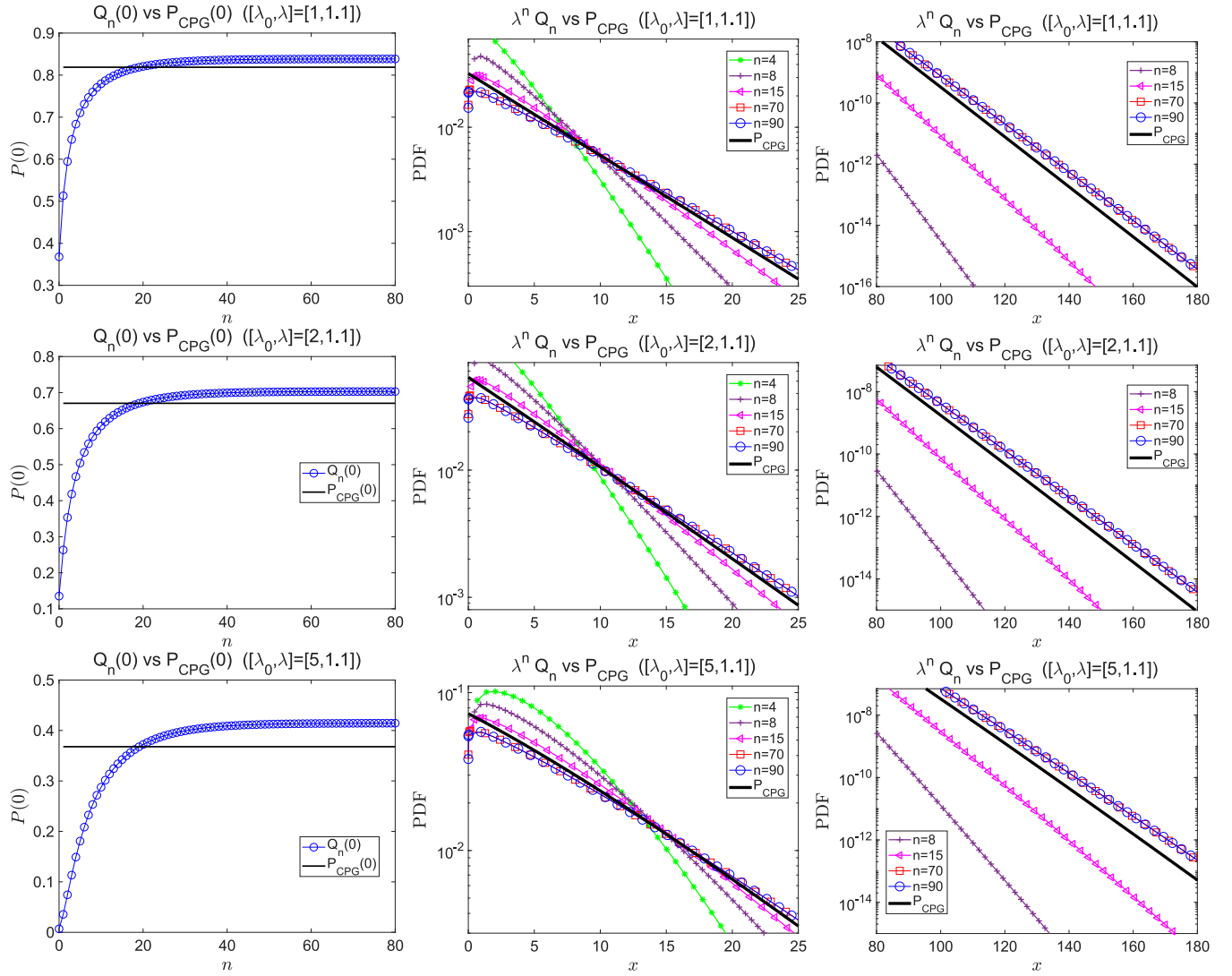


FIG. 3. Comparison between P_{CPG} and Q_n in the Poisson case under Condition II at $\lambda = 1.1$. Three different values of λ_0 , 1, 2, 5 are examined from top to bottom. The left columns are for the plot of $Q_n(0)$ (joined blue circles) against n , compared with $P_{\text{CPG}}(0)$ (black line). The other columns compare $\lambda^n Q_n(\lambda^n x)$ and $P_{\text{CPG}}(x)$ in the semilogarithmic scale. The bulk region and the right tail of the distribution are exhibited in the center and right columns, respectively.

discussion). Another intriguing observation, as shown in the bottom panels, is that for large λ , $P_{\text{CPG}}(x)$ no longer approximates $\lambda^n P_n(\lambda^n x)$ at all. This implies that, although the CPG distribution remains a relatively good approximation for the Poisson offspring distribution even for large λ , the same does not hold for the geometric offspring distribution.

C. Results under Condition II

We conducted numerical experiments under Condition II, assuming that Z_0 and $X_n^{(i)}$ in Eq. (1) follow Poisson distributions with means λ_0 and λ , respectively. Although this setting is the simplest, it remains nontrivial; it has been adopted as a reasonable model for SER distributions in earlier studies [10,32]. Note that analogous calculations for the geometric offspring distribution can be performed by appropriately modifying the procedure described in Appendix C.

We let $Q_n(\ell)$ denote the probability $\mathbb{P}(Z_n = \ell)$ under Condition II. We first examined cases where the mean λ of the offspring distribution is close to unity. Figure 3 shows results for $\lambda_0 = 1, 2, 5$ at $\lambda = 1.1$. As in Figs. 1 and 2, both the probability of zero (left panels) and the bulk region of $P_{\text{CPG}}(x)$ (center panels) are consistent with $\lambda^n Q_n(\lambda^n x)$ for all values of λ_0 examined, again supporting our theoretical analysis.

To further investigate the dependence on λ , we also examined cases with larger values of λ . Figure 4 shows the plot of $Q_n(0)$ against n and of $\lambda^n Q_n(\lambda^n x)$ in the top and bottom rows, respectively: The value of λ_0 was fixed to unity, and the results with $\lambda = 1.5, 2, 5$ are shown from left to right. As λ deviates from 1, the discrepancy between Q_n and P_{CPG} increases. An interesting observation in contrast to Fig. 1 is that the CPG distribution with the theoretical parameter values (45) is not a good approximation for large λ , even though the offspring distribution is Poisson. This observation and the

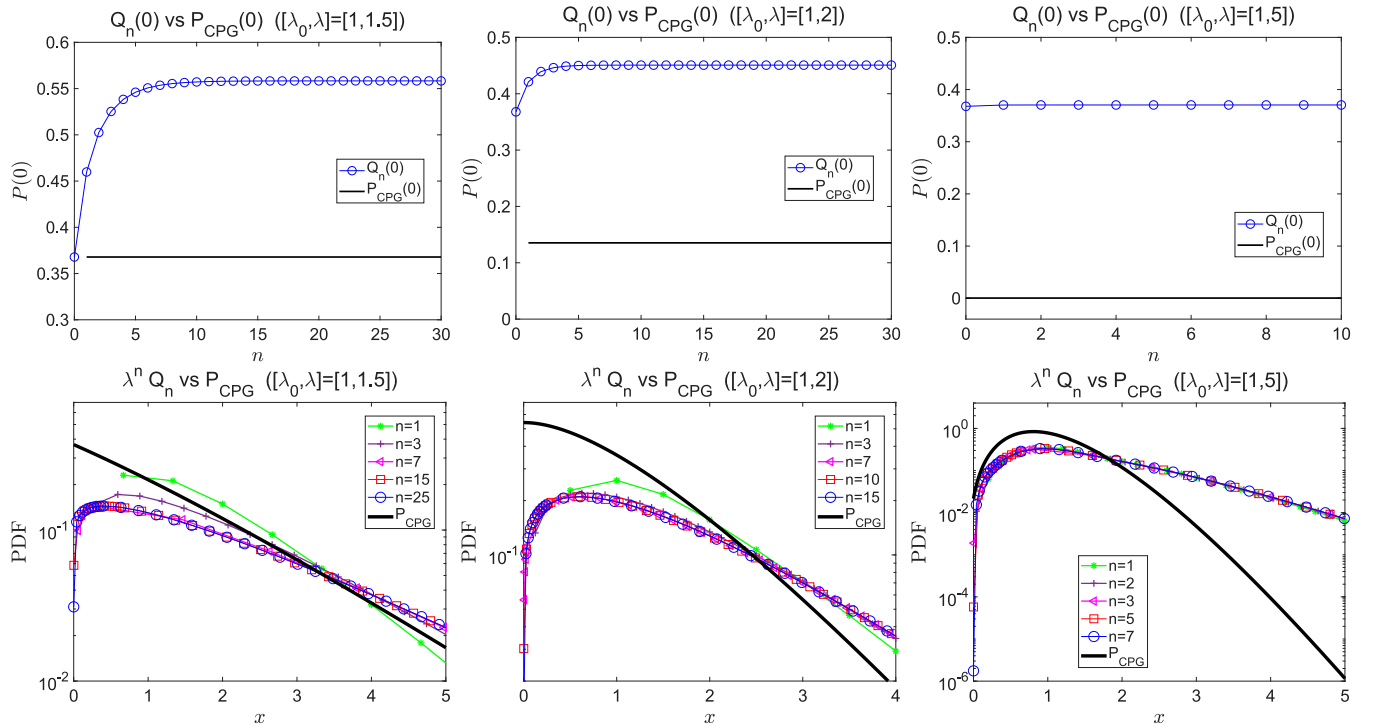


FIG. 4. Comparison between P_{CPG} and Q_n in the Poisson case at $\lambda_0 = 1$. Three different values of λ , 1.5, 2, 5, are examined from left to right. The top row gives the plot of $Q_n(0)$ (joined blue circles) against n , compared with $P_{\text{CPG}}(0)$ (black line). The bottom row compares $\lambda^n Q_n(\ell = \lambda^n x)$ (joined color markers) and $P_{\text{CPG}}(x)$ (black line) in the semilogarithmic scale.

one in Fig. 2 imply that the CPG distribution becomes a good approximation for large λ only in rather limited cases.

D. Fitting CPG in the Poisson case

As observed in Figs. 1 and 4, the limiting distribution of the Poisson GW process seems to have a shape similar to the CPG distribution, even when λ is not close to 1. This observation suggests that, by appropriately tuning its parameters, the CPG distribution may still work as a good approximation for the limiting distribution even when λ significantly deviates from unity. Here, we investigate this possibility.

To this end, we fit the CPG distribution to the numerically computed Q_n at large n and investigate how well the CPG distribution can fit Q_n . The fitting procedure we adopted consists of the following two steps:

- (1) Estimate the parameter μ on the basis of the probability $Q_n(0)$ only, that is, $\hat{\mu} = -\ln Q_n(0)$.
- (2) Perform least-squares fitting of $P_{\text{CPG}}(\ell/\lambda^n; \hat{\mu}, \alpha, \tau)$ to $\lambda^n Q_n(\ell)$ with respect to (α, τ) over an appropriate range of $\ell(> 0)$, obtaining $(\hat{\alpha}, \hat{\tau})$.

The fitting results are shown in Fig. 5. The left panel corresponds to the bottom right panel of Fig. 4 in the parameters but compares the fitted CPG (black line) and Q_n for $n = 7$ (red line with squares), as well as the CPG of the parameters with the theoretical values (45) (blue dashed line). The fitted result showed a nice agreement with Q_n over a wide probability region, though there was a gap in the tail with small probabilities. This suggests that while the CPG distribution is indeed an approximation and does not precisely describe the limiting

distribution, it is still useful for approximating regions with high-probability values. For further examination, the cases $(\lambda_0, \lambda) = (12, 2.7)$ and $(100, 2)$ are shown in the middle and right panels, respectively: The former set of the parameter values was taken from the example values of dynode gain of EM in Ref. [10], while the latter one was intended for examining scenarios where λ_0 is extremely large. In these two cases, the fitted CPGs showed excellent agreement with Q_n at large n even at very small probability values, suggesting that the CPG approximation improves as λ_0 increases. However, this result may not be entirely surprising because the distributions of both the GW and CPG models approach a normal distribution as λ_0 and $\hat{\mu}$ increase (Ref. [33], Theorem 3.7.1). The theoretical and fitted values of the CPG parameters for these three cases are presented in Table I.

Overall, the observation in this section supports the use of the CPG distribution to describe the limiting distribution of the Poisson GW process even when λ is not close to unity. This result suggests the potential of the CPG distribution to model data distributions generated from supercritical GW processes with any λ in practical applications.

TABLE I. The theoretical and fitted values of CPG parameters.

	Theory $(\hat{\mu}, \hat{\alpha}, \hat{\tau})$	Fitted $(\hat{\mu}, \hat{\alpha}, \hat{\tau})$
$(\lambda_0, \lambda) = (1, 5)$	(8, 1, 0.125)	(0.99, 3.80, 0.275)
$(\lambda_0, \lambda) = (12, 2.7)$	(40.8, 1, 0.294)	(11.0, 2.19, 0.500)
$(\lambda_0, \lambda) = (100, 2)$	(200, 1, 0.5)	(79.7, 1.68, 0.745)

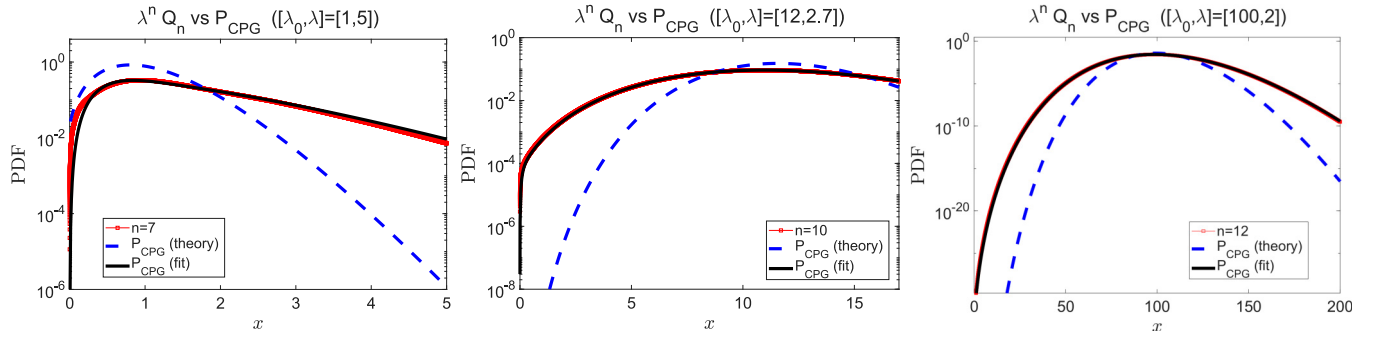


FIG. 5. Fitted P_{CPG} (black line), Q_n for large n (joined red squares), and P_{CPG} with the theoretical parameter values (blue dashed line) are compared in the Poisson case with λ fairly larger than unity. Three different parameter sets, $(\lambda_0, \lambda) = (1, 5)$, $(12, 2.7)$, and $(100, 2)$ from left to right, are considered. In all the cases, the fitted CPG distribution recovers a nice agreement with Q_n .

IV. DISCUSSION

A. Tail behaviors

In general, if the moment generating function $M_X(t) = \mathbb{E}(e^{tX})$, or the cumulant generating function $\ln M_X(t)$, of a random variable X diverges at $t = t_0 > 0$, it implies that the tail probability $\mathbb{P}(X > x)$ of X behaves as $\mathbb{P}(X > x) \sim C e^{-t_0 x}$ with some $C > 0$ as $x \rightarrow \infty$. In other words, the distribution of X has an exponential tail. A precise statement is given in the next proposition.

Proposition 1. Let X be a nonnegative random variable, $F_X(x) := \mathbb{P}(X \leq x)$ be the cumulative distribution function of X , and $M_X(t) = \mathbb{E}(e^{tX})$ be the moment generating function of X . Let $\bar{F}_X(x) := 1 - F_X(x)$. Then,

- (1) Assume $M_X(t) < \infty$ for some $t > 0$. One then has $\lim_{x \rightarrow \infty} e^{tx} \bar{F}_X(x) = 0$.
- (2) Assume $\sup_x e^{tx} \bar{F}_X(x) < \infty$ for some $t > 0$. Then for any $t' \in [0, t)$, one has $M_X(t') < \infty$.

A proof of Proposition 1 is given in Appendix D.

Tail behavior of the distribution of the limiting random variable W is linked to that of the distribution of the offspring distribution, as shown in the following proposition.

Proposition 2 (Proposition 3 in [34]). Let $f(s)$ and $M_W(t)$ be the probability generating function of the offspring distribution and the moment generating function of W . Let

$$s_0 := \sup\{s \geq 0 : f(s) < \infty\}, \quad (49)$$

$$t_0 := \sup\{t : M_W(t) < \infty\}. \quad (50)$$

One then has

- (1) $s_0 = \infty$ if and only if $t_0 = \infty$.
- (2) $s_0 \in (1, \infty)$ if and only if $t_0 \in (0, \infty)$.

Propositions 1 and 2 together imply that the distribution of the limiting random variable W has an exponential tail if the offspring distribution has an exponential tail, as well as that the former has a lighter-than-exponential tail if the latter has a lighter-than-exponential tail.

Returning to our asymptotic results, one observes that the moment generating functions in both Conditions I and II diverge at $t = 2(\lambda - 1)/\kappa_2^* < \infty$, implying that the corresponding distribution has an exponential tail regardless of the offspring distribution. This is in conflict with Proposition 2, showing that our asymptotic result does not in general reproduce the true tail behavior of W .

B. Diffusion approximation

Diffusion approximation of the GW processes [35–38] is closely related to our framework. The diffusion approximation of the GW process is obtained via rescaling of the state space and time. More concretely, one counts generation n as well as the population size Z_n as multiples of $\Delta \gg 1$, via

$$Y_s = \frac{Z_s \Delta}{\Delta}, \quad (51)$$

and studies the diffusion process defined in the limit $\Delta \rightarrow \infty$, which is called the diffusion branching process.

Assume a family of the diffusion branching processes indexed by $\Delta > 0$. For the offspring distribution, we assume $\kappa_1 = \lambda = 1 + \alpha/\Delta$ with $\alpha \in (-\infty, \infty)$, and $\kappa_2^* := \lim_{\Delta \rightarrow \infty} \kappa_2 \in (0, \infty)$. We also assume $Y_0 = z_0 \Delta$ with $z_0 \in [0, \infty)$. The Feller-Jirina theorem [35,36] (see also Ref. [38]) states that, under some regularity conditions, the cumulant generating function $\xi(t)$ of Y_s in the limit $\Delta \rightarrow \infty$ is given by

$$\xi(t) = \frac{z_0 t e^{\alpha s}}{1 - \frac{\kappa_2^* u_s(\alpha)}{2} t}, \quad (52)$$

where

$$u_s(\alpha) = \begin{cases} \frac{e^{\alpha s} - 1}{\alpha}, & \alpha \neq 0, \\ s, & \alpha = 0. \end{cases} \quad (53)$$

One can observe that the state distribution of the diffusion branching process in the criticality limit is again given by a CPG distribution.

It should be noted that in deriving the above result one takes the limit $\Delta \rightarrow \infty$ while $(\lambda - 1)n = \frac{\alpha}{\Delta} \cdot s \Delta = \alpha s$ is kept finite. In other words, one takes the limits $\lambda \rightarrow 1$ and $n \rightarrow \infty$ simultaneously. In our argument, on the other hand, we take the limit $n \rightarrow \infty$ first, and then consider the asymptotic $\lambda \downarrow 1$. It is interesting to observe that, in spite of the difference in how to take the limits, one arrives at the same CPG distribution. It should also be noted that the above result states that the distribution of Y_s in the limit $\Delta \rightarrow \infty$ has an exponential tail, just as our asymptotic result. As argued in Sec. IV A, it may not be consistent with the limiting distribution of W , which has a lighter-than-exponential tail if the offspring distribution has a lighter-than-exponential tail.

C. Large- n asymptotic

Garcia-Millan *et al.* [39] performed an analysis on the finite-size scaling of the survival probability $\mathbb{P}(Z_n > 0)$ of the GW branching process by considering sufficiently large but finite $n = m + \ell$ with $\ell \gg m \gg 1$. One can extend their analysis to discuss the large- n asymptotic of the cumulant generating function $K_{\bar{W}_n}(t)$ of $\bar{W}_n = (\lambda - 1)Z_n/\lambda^n$ as $\lambda \downarrow 1$.

For the analysis below, we impose the following assumptions on the offspring distribution.

Assumption 1.

(1) The offspring distribution is parametrized by its mean λ .

(2) The cumulant generating function

$$\psi(t) = \sum_{v=1}^{\infty} \frac{\kappa_v}{v!} t^v \quad (54)$$

of the offspring distribution exists, where κ_v denotes the v th cumulant of the offspring distribution and $\kappa_1 = \lambda$.

(3) $\lim_{\lambda \downarrow 1} \kappa_2 = \kappa_2^* \in (0, \infty)$.

(4) For any $v \geq 3$, the order of the v th cumulant κ_v is $o(1/(\lambda - 1)^{v-2})$ in the limit $\lambda \downarrow 1$.

Several common discrete probability distributions, including the Poisson and negative binomial distributions, satisfy these assumptions. It should also be noted that these assumptions are weaker than those in Sec. II A.

Under Assumption 1 and Condition I, $K_{\bar{W}_m}(t)$ satisfies the functional equation

$$K_{\bar{W}_{m+1}}(\lambda t) = \psi(K_{\bar{W}_m}(t)), \quad (55)$$

where $\psi(t)$ denotes the cumulant generating function of the offspring distribution. Near the criticality, we may assume that, for sufficiently large m , $K_{\bar{W}_m}(t)$ is close to $(\lambda - 1)t/(1 - \frac{\kappa_2^*}{2}t)$, which is $\mathcal{O}(\lambda - 1)$ as $\lambda \downarrow 1$. Under this assumption, one may take account of terms up to second order in $(\lambda - 1)$ while ignoring higher-order terms in Eq. (55), yielding

$$\begin{aligned} \frac{1}{K_{\bar{W}_{m+1}}(\lambda t)} &= \frac{1}{\psi(K_{\bar{W}_m}(t))} \approx \frac{1}{\lambda K_{\bar{W}_m}(t) + \frac{\kappa_2^*}{2} K_{\bar{W}_m}(t)^2} \\ &= \frac{1}{\lambda K_{\bar{W}_m}(t) \left(1 + \frac{\kappa_2^* K_{\bar{W}_m}(t)}{2\lambda}\right)} \\ &\approx \frac{1}{\lambda K_{\bar{W}_m}(t)} \left(1 - \frac{\kappa_2^* K_{\bar{W}_m}(t)}{2\lambda}\right) \\ &= \frac{1}{\lambda K_{\bar{W}_m}(t)} - \frac{\kappa_2^*}{2\lambda^2}. \end{aligned} \quad (56)$$

Note that, in both the second and fourth lines, the approximations are performed by noting that $K_{\bar{W}_m}(t) = \mathcal{O}(\lambda - 1)$ is a small quantity as $\lambda \downarrow 1$ and by retaining terms up to the second order in $(\lambda - 1)$.

Since

$$\frac{1}{K_{\bar{W}_m}(t)} = \frac{1 - \frac{\kappa_2^*}{2\lambda}t}{(\lambda - 1)t} \quad (57)$$

is the fixed point of Eq. (56), we let

$$h_m(t) := \frac{1}{K_{\bar{W}_m}(t)} - \frac{1 - \frac{\kappa_2^*}{2\lambda}t}{(\lambda - 1)t} \quad (58)$$

be the deviation of $1/K_{\bar{W}_m}(t)$ from the fixed point. The above equation can then be rewritten in terms of h_m as the following simple formula:

$$h_{m+1}(\lambda t) = \frac{1}{\lambda} h_m(t). \quad (59)$$

Let $\mathcal{F}$ be the operator that maps $h(t)$ to $\frac{1}{\lambda}h(\frac{t}{\lambda})$. The above equation can then be represented concisely as $h_{m+1} = \mathcal{F}(h_m)$. It should be noted that $\mathcal{F}$ is a linear operator, which can be confirmed via

$$\begin{aligned} \mathcal{F}((af + bg)(t)) &= \frac{1}{\lambda}(af + bg)\left(\frac{t}{\lambda}\right) \\ &= \frac{a}{\lambda}f\left(\frac{t}{\lambda}\right) + \frac{b}{\lambda}g\left(\frac{t}{\lambda}\right) \\ &= a\mathcal{F}(f(t)) + b\mathcal{F}(g(t)). \end{aligned} \quad (60)$$

One also has

$$\mathcal{F}(t^k) = \frac{1}{\lambda} \left(\frac{t}{\lambda}\right)^k = \frac{t^k}{\lambda^{k+1}}, \quad (61)$$

showing that t^k is the eigenfunction of $\mathcal{F}$ associated with the eigenvalue $1/\lambda^{k+1}$.

It is known that $\bar{W}_m$ has mean $\lambda - 1 > 0$, which allows us to apply Wronski's formula (Ref. [40], Theorem 1.3) to a power series expansion of $K_{\bar{W}_m}(t)/t$. One consequently finds that $h_m(t)$ admits the following series expansion:

$$h_m(t) = \sum_{k=0}^{\infty} c_k t^{k-1}. \quad (62)$$

Repeated application of the operator $\mathcal{F}$ to h_m yields

$$h_{m+\ell}(t) = \mathcal{F}^\ell(h_m) = \sum_{k=0}^{\infty} \frac{c_k}{\lambda^{\ell k}} t^{k-1}. \quad (63)$$

This in turn can be translated into the following formula for $K_{\bar{W}_{m+\ell}}(t)$:

$$K_{\bar{W}_{m+\ell}}(t) = \frac{1}{h_{m+\ell}(t) + \frac{1 - \frac{\kappa_2^*}{2\lambda}t}{(\lambda - 1)t}} = \frac{(\lambda - 1)t}{1 + c_0(\lambda - 1) - \left(\frac{\kappa_2^*}{2\lambda} - \frac{c_1(\lambda - 1)}{\lambda^\ell}\right)t + \sum_{k=2}^{\infty} \frac{c_k(\lambda - 1)}{\lambda^{\ell k}} t^k}. \quad (64)$$

The coefficients c_0 and c_1 are given by

$$c_0 = 0, \quad (65)$$

TABLE II. Asymptotic and limiting behaviors of GW branching processes near criticality. Our results are colored in blue. “ $X \sim \text{Dist}$ ” means that random variable X follows the distribution Dist . “ $X | Y > 0 \sim \text{Dist}$ ” means that random variable X follows the distribution Dist conditionally on the event $Y > 0$.

λ	n large but finite	$n = \infty$
Supercritical $\lambda \downarrow 1$	$\frac{Z_n}{\lambda^n} \sim \text{CPG}$ (this paper, Sec. IV C)	Explosion: $\lim_{n \rightarrow \infty} Z_n Z_n > 0 = \infty$ [13, Theorem 6.2] $\lim_{n \rightarrow \infty} \frac{Z_n}{\lambda^n} \sim \text{CPG}$ (this paper, Sec. II)
Critical $\lambda = 1$	$Z_n Z_n > 0 \sim \text{Geometric}$ (Ref. [25], Sec. III D)	Extinction: $\lim_{n \rightarrow \infty} Z_n = 0$ [13, Theorem 6.1] $\lim_{n \rightarrow \infty} \frac{Z_n}{n} Z_n > 0 \sim \text{Exponential}$ [13, Theorem 10.1]
Subcritical $\lambda \uparrow 1$		Extinction: $\lim_{n \rightarrow \infty} Z_n = 0$ [13, Theorem 6.1] $\lim_{n \rightarrow \infty} Z_n Z_n > 0 \sim \text{Geometric}$ (Ref. [25], Sec. III D)

$$c_1 = \frac{\kappa_2^*}{2\lambda(\lambda - 1)} - \frac{\kappa_2 \sum_{i=1}^m \lambda^{i-1}}{2\lambda^{m+1}}, \quad (66)$$

respectively (see Appendix E for the derivation). Therefore, Eq. (64) can be rewritten as

$$K_{\tilde{W}_{m+\ell}}(t) = \frac{(\lambda - 1)t}{1 - \left\{ \frac{\kappa_2^*(1-\lambda^{-\ell})}{2\lambda} + \frac{\kappa_2 \sum_{i=1}^m \lambda^{i-1}}{2\lambda^{m+\ell+1}} (\lambda - 1) \right\} t + \sum_{k=2}^{\infty} \frac{c_k(\lambda-1)}{\lambda^k} t^k}. \quad (67)$$

We now fix m and assume the asymptotic of ℓ tending large, via the scaling defined by $\ell = s\Delta$ and $\lambda - 1 = \alpha/\Delta$ in the limit $\Delta \rightarrow \infty$. This scaling yields $\lambda = 1 + \alpha/\Delta$ and $\lambda^\ell \approx e^{\alpha s}$. Moreover, the order of c_k for $k \geq 2$ is $o(1)$ in the asymptotic regime $\lambda \downarrow 1$ (see Appendix E for a detailed discussion). One therefore obtains

$$K_{\tilde{W}_n}(t) = K_{\tilde{W}_{m+\ell}}(t) \approx \frac{\alpha t / \Delta}{1 - \frac{\kappa_2^*}{2}(1 - e^{-\alpha s})t}. \quad (68)$$

This result suggests that when λ is larger than, but close enough to, 1, the distribution of $\tilde{W}_n$ for large $n = s\Delta$ will be well approximated by the CPG distribution whose cumulant generating function is given by

$$\tilde{K}(t) = \frac{\alpha t / \Delta}{1 - \frac{\kappa_2^*}{2}(1 - e^{-\alpha s})t}. \quad (69)$$

The cumulant generating function $\xi(t)$ of $Y_s := Z_{s\Delta}/\Delta$ with the initial condition $Y_0 = z_0\Delta$ and in the limit $\Delta \rightarrow \infty$ is derived from the above result via noting that

$$Y_s = \frac{Z_{s\Delta}}{\Delta} = \frac{1}{\Delta} \frac{\lambda^{s\Delta}}{\lambda - 1} \tilde{W}_{s\Delta} = \frac{1}{\Delta} \frac{(1 + \alpha/\Delta)^{s\Delta}}{\alpha/\Delta} \tilde{W}_{s\Delta} \\ \xrightarrow{\Delta \rightarrow \infty} \frac{e^{\alpha s}}{\alpha} \tilde{W}_{s\Delta}, \quad (70)$$

as

$$\xi(t) = \frac{z_0 t e^{\alpha s}}{1 - \frac{\kappa_2^*(e^{\alpha s} - 1)}{2\alpha} t}. \quad (71)$$

It should be noted that the above result is identical to formula (52) in the Feller-Jirřna theorem for the diffusion branching processes.

Table II summarizes known results, as well as our contributions, about asymptotic and limiting behaviors of the GW process in the criticality limit $\lambda \rightarrow 1$. Our results fill

the remaining theoretical gap for the supercritical case, complementing the existing results for the critical and subcritical cases, and thus provide a more comprehensive picture of GW branching processes near criticality.

V. CONCLUSION

In this work, we studied the asymptotics of supercritical GW processes as they approach criticality, $\lambda \downarrow 1$. We derived a probability distribution in closed form whose cumulant generating function matches that of the limiting distribution of the GW process up to the leading order in $\lambda - 1$ and showed that the derived distribution belongs to a class of CPG distributions. Numerical experiments revealed that these CPG models were in good agreement with the corresponding GW models for sufficiently large generations under a reasonable parameter regime. We also showed that, even when λ is not close to unity and the offspring distribution is Poisson, the appropriately fitted CPG model closely matched the corresponding GW model.

In the discussion section, we argued the tail behaviors of the distribution of the limiting random variable of the supercritical GW process. While this discussion implies the limitation of the approximation capability of the CPG model, it should be noted that this limitation may not pose a significant problem in practical applications: For example, one is typically interested more in bulk behaviors rather than in tail behaviors in the analysis of data from EMs. We furthermore provided discussion to clarify the position and contribution of our work in comparison with other related studies. They together suggest that the CPG distribution would provide a good approximation of the population-size distribution of the GW processes in the asymptotic $\lambda \downarrow 1$.

Our results suggest the potential of the CPG model to analyze datasets generated from cascaded multiplication pro-

cesses, such as detection signals of EMs and population sizes of individuals with specific biological characteristics, offering various mathematical and statistical advantages for data analysis. In particular, its application holds promise for large-scale datasets, which have been challenging to analyze using supercritical GW models.

ACKNOWLEDGMENTS

This work was partially supported by JST, CREST Grant No. JPMJCR1862, Japan (K.U., T.O., T.T.), JSPS KAKENHI under Grants No. 22K12179 (T.O.), No. 21J23154 (K.U.), Grant-in-Aid for Transformative Research Areas (A), “Foundation of machine learning physics” (Grant No. 22H05117) (T.O.), and Grant-in-Aid for Transformative Research Areas (A), “Shin-biology regulated by protein lifetime” (Grant No. 24H01895) (T.T.).

DATA AVAILABILITY

The data that support the findings of this article are openly available [41].

APPENDIX A: RECURSION UNDER CONDITION I

In this Appendix, we derive the extended recursion formula for the probability distribution of Z_n under Condition I. We write the probability distribution of Z_n under Condition I as

$$P_n(\ell) \equiv \mathbb{P}(Z_n = \ell) \quad (n \geq 0), \quad (\text{A1})$$

where $P_0(\ell) = \mathbb{P}(Z_0 = \ell) = \delta_{\ell,1}$ by definition, where $\delta_{i,j}$ denotes Kronecker’s delta, with value 1 if $i = j$ and 0 otherwise. The key to deriving the formula is the recursive structure (6) of the probability generating functions, which reads

$$f_n(s) = f(f_{n-1}(s)). \quad (\text{A2})$$

Evaluating the ℓ th derivative of both sides of Eq. (A2) yields

$$\frac{d^\ell}{ds^\ell} f_n(s) = \frac{d^{\ell-1}}{ds^{\ell-1}} f'(f_{n-1}(s)) f'_{n-1}(s). \quad (\text{A3})$$

When the offspring distribution is Poisson with mean λ , one has $f'(s) = \lambda f(s)$, on the basis of which the equality $f'(f_{n-1}(s)) = \lambda f(f_{n-1}(s)) = \lambda f_n(s)$ holds, where the second equality is due to Eq. (A2) [9]. One can then represent the right-hand side of Eq. (A3) as a polynomial of the derivatives of $f_n f'_{n-1}$ with simple coefficients. Since the probability distribution is derived from the probability generating function via

$$P_n(\ell) = \frac{1}{\ell!} \left. \frac{d^\ell}{ds^\ell} f_n(s) \right|_{s=0}, \quad (\text{A4})$$

the derivatives of f_n and f'_{n-1} are converted to the probability distributions through Eq. (A4), which allows us to convert Eq. (A3) into a recursion relation between P_n and P_{n-1} .

We extend the above-mentioned derivation in Ref. [9] via assuming the condition $f'(s) = a(f(s))^m$ for the probability generating function f of the offspring distribution, with some a and a positive integer m . Noting that $f(1) = 1$ and $f'(1) = \lambda$ hold, one should have $a = \lambda$, so that our assumption now

reads

$$f'(s) = \lambda(f(s))^m. \quad (\text{A5})$$

As mentioned above, the Poisson offspring distribution satisfies this assumption with $m = 1$, which is the case dealt with in Ref. [9]. The geometric offspring distribution satisfies it with $m = 2$. More generally, for $m \in \{2, 3, \dots\}$, the offspring distribution, which is negative binomial with parameters $r = 1/(m-1)$ and $p \in (0, 1)$ satisfies it with $\lambda = r(1-p)/p$. We show this fact.

The probability function of the negative binomial distribution [42] with the parameter $r > 0$ and the mean $\lambda = r(1-p)/p$ is given by

$$P_{\text{NB}}(x) = \binom{x+r-1}{x} p^r (1-p)^x, \quad x \in \{0, 1, 2, \dots\}. \quad (\text{A6})$$

Its probability generating function is given by

$$f_{\text{NB}}(s) = \left(\frac{r}{r + \lambda - \lambda s} \right)^r. \quad (\text{A7})$$

Here, the parameter $p \in (0, 1]$ is usually called the success probability and it is related to the mean λ as

$$p = \frac{r}{r + \lambda}. \quad (\text{A8})$$

Letting $r = 1$ in the negative binomial distribution yields the geometric distribution

$$P_{\text{geo}}(x) = p(1-p)^x, \quad x \in \{0, 1, 2, \dots\}. \quad (\text{A9})$$

It is readily seen that f_{NB} with $m = 1 + 1/r \in \mathbb{N}$ satisfies Eq. (A5).

The assumption (A5) allows us to rewrite the right-hand side of Eq. (A3) as

$$\begin{aligned} & \frac{d^{\ell-1}}{ds^{\ell-1}} f'(f_{n-1}(s)) f'_{n-1}(s) \\ &= \lambda \frac{d^{\ell-1}}{ds^{\ell-1}} (f_n(s))^m f'_{n-1}(s) \\ &= \lambda \sum_{i_1+i_2+\dots+i_{m+1}=\ell-1} \binom{\ell-1}{i_1, i_2, \dots, i_{m+1}} \\ & \quad \times f_n^{(i_1)}(s) \cdots f_n^{(i_m)}(s) f_{n-1}^{(i_{m+1}+1)}(s), \end{aligned} \quad (\text{A10})$$

where the first equality is obtained by combining Eqs. (A2) and (A5), and where the last line is derived using the generalized Leibniz rule: $f^{(i)}$ denotes the i th derivative of f , and $\sum_{i_1+i_2+\dots+i_{m+1}=\ell-1}$ denotes the summation over the set $\{(i_1, i_2, \dots, i_{m+1}) \in \{0, 1, \dots, \ell-1\}^{m+1} \mid i_1 + i_2 + \dots + i_{m+1} = \ell-1\}$ of multi-indices. Inserting this into Eq. (A3), dividing both sides by $\ell!$, and substituting $s = 0$, we arrive at the following recursion:

$$\begin{aligned} P_n(\ell) &= \frac{\lambda}{\ell} \sum_{i_1+i_2+\dots+i_{m+1}=\ell-1} (i_{m+1} + 1) \\ & \quad \times P_n(i_1) \cdots P_n(i_m) P_{n-1}(i_{m+1} + 1), \end{aligned} \quad (\text{A11})$$

which is valid for any $n \geq 2$. For $n = 1$, $P_1(\ell) = \mathbb{P}(Z_1 = \ell) = \mathbb{P}(X_1 = \ell)$ is nothing but the offspring distribution. The

boundary condition $P_n(0)$ necessary for Eq. (A11) is obtained from the recursion as well:

$$P_n(0) = f_n(0) = f(f_{n-1}(0)) = f(P_{n-1}(0)). \quad (\text{A12})$$

By recursively evaluating Eqs. (A11) and (A12), $(P_n)_n$ is determined without ambiguity. Since the offspring distribution that we assume (Poisson or negative binomial) is with infinite support, one needs to truncate the range of ℓ in calculating $(P_n)_n$.

APPENDIX B: RECURSION UNDER CONDITION II

In this Appendix, we derive the extended recursion formula for the probability distribution of Z_n under Condition II. We write the probability distribution of Z_n under Condition II as

$$Q_n(\ell) \equiv \mathbb{P}(Z_n = \ell) \quad (n \geq 0). \quad (\text{B1})$$

The corresponding probability generating function is denoted as $g_n(s)$ hereafter. Using this notation and the underlying GW process (1), the following recurrence relation can be derived:

$$g_n(s) = g_0 \circ \overbrace{(f \circ f \circ \dots \circ f)}^{n \text{ times}}(s) = g_0(f_n(s)) \quad (n \geq 0), \quad (\text{B2})$$

where $f(s)$ denotes the probability generating function of $X_n^{(i)}$ and where $f_n(s)$ is the probability generating function of Z_n conditional on $Z_0 = 1$ as in Condition I. Evaluating the ℓ th derivative of both sides of Eq. (B2) leads to

$$\frac{d^\ell}{ds^\ell} g_n(s) = \frac{d^{\ell-1}}{ds^{\ell-1}} g'_0(f_n(s)) f'_n(s). \quad (\text{B3})$$

Similarly to (A5), for the probability generating function g_0 of Z_0 we assume

$$g'_0(s) = \lambda_0 (g_0(s))^q, \quad (\text{B4})$$

with a positive integer q , where $\lambda_0 > 0$ is the mean of Z_0 . It amounts to assuming that Z_0 is either Poisson ($q = 1$) or negative binomial ($q \in \{2, 3, \dots\}$). The case $q = m = 1$ was dealt with in Ref. [10]. Equation (B3) is thus rewritten as

$$\begin{aligned} \frac{d^\ell}{ds^{\ell-1}} g'_0(f_n(s)) f'_n(s) \\ = \lambda_0 \sum_{i_1+i_2+\dots+i_{q+1}=\ell-1} \binom{\ell-1}{i_1, i_2, \dots, i_{q+1}} \\ \times g_n^{(i_1)}(s) \dots g_n^{(i_q)}(s) f_n^{(i_{q+1}+1)}(s). \end{aligned} \quad (\text{B5})$$

Inserting this into Eq. (B3) and dividing both sides by $\ell!$, we obtain

$$\begin{aligned} Q_n(\ell) = \frac{\lambda_0}{\ell} \sum_{i_1+i_2+\dots+i_{q+1}=\ell-1} (i_{q+1} + 1) \\ \times Q_n(i_1) \dots Q_n(i_q) P_n(i_{q+1} + 1), \end{aligned} \quad (\text{B6})$$

at $s = 0$. This formula holds for $n \geq 1$ and the boundary condition $Q_n(0)$ is given by

$$Q_n(0) = g_0(f_n(0)) = g_0(P_n(0)). \quad (\text{B7})$$

The solution to Eqs. (B6) and (B7) is efficiently computed recursively on top of $(P_n)_n$, which are evaluated with Eqs.

(A11) and (A12). These constitute the algorithmic basis for computing the distributions of $(Z_n)_n$ under Condition II.

APPENDIX C: RECURSION SPECIFIC TO THE GEOMETRIC DISTRIBUTION

The probability generating function of the geometric distribution is given by Eq. (A7) with $r = 1$. Combining this with Eq. (A2), it is easy to show that f_n has a linear fractional form and that its parameters satisfy the following recursive equation:

$$f_n(s) = \frac{c_n + d_n s}{a_n - b_n s}, \quad (\text{C1})$$

$$\begin{pmatrix} a_{n+1} \\ b_{n+1} \\ c_{n+1} \\ d_{n+1} \end{pmatrix} = \begin{pmatrix} (1+\lambda)a_n - \lambda c_n \\ (1+\lambda)b_n + \lambda d_n \\ a_n \\ -b_n \end{pmatrix}, \quad (\text{C2})$$

with the initial condition $(a_1, b_1, c_1, d_1) = (1 + \lambda, \lambda, 1, 0)$. By solving this numerically, the probability generating function for any n can be easily computed. Note that directly iterating Eq. (C2) causes the parameter values to grow rapidly, and hence it is technically better to normalize them by multiplying an appropriate identical factor to both the numerator and denominator of Eq. (C1). Given f_n , the distribution of Z_n can be computed by a power series expansion of the denominator of Eq. (C1), yielding

$$P_n(0) = \frac{c_n}{a_n}, \quad P_n(\ell) = \frac{b_n^{\ell-1}}{a_n^{\ell+1}} (c_n b_n + a_n d_n) \quad (\ell \geq 1). \quad (\text{C3})$$

APPENDIX D: PROOF OF PROPOSITION 1

Proof. We first prove (a). Assume that $M_X(t) = \mathbb{E}(e^{tX})$ is finite at some $t > 0$. From the Chernoff bound, one has, for any $x \geq 0$,

$$\begin{aligned} \bar{F}_X(x) = \mathbb{P}(X > x) &\leq \mathbb{P}(X \geq x) = \mathbb{P}(e^{tX} \geq e^{tx}) \\ &\leq \frac{\mathbb{E}(e^{tX})}{e^{tx}} = \frac{M_X(t)}{e^{tx}}. \end{aligned} \quad (\text{D1})$$

Letting $\tau_X(t, x) = e^{tx} \bar{F}_X(x)$, the above inequality means that $\tau_X(t, x) \leq M_X(t) < \infty$ for any $x \geq 0$.

For $z \geq 0$, let

$$m_X(t, z) := \int_0^z e^{tx} dF_X(x) = - \int_0^z e^{tx} d\bar{F}_X(x). \quad (\text{D2})$$

One then has $M_X(t) = \lim_{z \rightarrow \infty} m_X(t, z)$. The integration-by-parts formula gives

$$\begin{aligned} m_X(t, z) &= -[e^{tx} \bar{F}_X(x)]_0^z + \int_0^z t e^{tx} \bar{F}_X(x) dx \\ &= -\tau_X(t, z) + \bar{F}_X(0) + t \int_0^z \tau_X(t, x) dx \end{aligned} \quad (\text{D3})$$

$$\geq -M_X(t) + \bar{F}_X(0) + t \int_0^z \tau_X(t, x) dx. \quad (\text{D4})$$

Taking the limit $z \rightarrow \infty$, one obtains

$$\int_0^\infty \tau_X(t, x) dx \leq \frac{2M_X(t) - \bar{F}_X(0)}{t} < \infty, \quad (\text{D5})$$

which, together with $\tau_X(t, x) \geq 0$, implies

$$\lim_{x \rightarrow \infty} \tau_X(t, x) = 0, \quad (\text{D6})$$

proving item (a).

We next prove (b). Let $h_X(t) := \sup_x \tau_X(t, x)$, which is finite by assumption. For any $z \in [0, \infty)$ and any $t' \in [0, t)$, one has

$$\begin{aligned} \int_0^z \tau_X(t', x) dx &= \int_0^z e^{-(t-t')x} \tau_X(t, x) dx \\ &\leq \int_0^z e^{-(t-t')x} h_X(t) dx \\ &= h_X(t) \frac{1 - e^{-(t-t')z}}{t - t'}. \end{aligned} \quad (\text{D7})$$

Combining this and Eq. (D3), one obtains

$$\begin{aligned} m_X(t', z) &\leq t h_X(t) \frac{1 - e^{-(t-t')z}}{t - t'} + \bar{F}_X(0) - \tau_X(t', z) \\ &\leq t h_X(t) \frac{1 - e^{-(t-t')z}}{t - t'} + \bar{F}_X(0). \end{aligned} \quad (\text{D8})$$

Taking the limit $z \rightarrow \infty$, one has

$$M_X(t') \leq \frac{t h_X(t)}{t - t'} + \bar{F}_X(0) < \infty, \quad (\text{D9})$$

proving item (b). $\blacksquare$

APPENDIX E: ORDER OF c_k IN THE ASYMPTOTIC $\lambda \downarrow 1$

The objective here is to evaluate the order of the coefficients $\{c_k\}$ in Eq. (62).

Lemma E1. Under Assumption 1, let

$$\bar{W}_m = \frac{(\lambda - 1)}{\lambda^m} Z_m, \quad (\text{E1})$$

where Z_m is a random variable defined by Eq. (1) with $Z_0 \equiv 1$. Assume that the cumulant generating function $K_{\bar{W}_m}(t)$ of $\bar{W}_m$ exists, i.e.,

$$K_{\bar{W}_m}(t) = \sum_{k=1}^{\infty} \frac{\bar{\kappa}_k^{(m)}}{k!} t^k \quad (\text{E2})$$

is absolutely convergent on an open interval containing $t = 0$. Then, for any m , the first and second cumulants of $\bar{W}_m$ are given by

$$\bar{\kappa}_1^{(m)} = \lambda - 1, \quad (\text{E3})$$

$$\bar{\kappa}_2^{(m)} = \frac{\kappa_2(\lambda - 1)}{\lambda} \left(1 - \frac{1}{\lambda^m}\right). \quad (\text{E4})$$

Moreover, for $k \geq 3$, the order of the k th cumulant $\bar{\kappa}_k^{(m)}$ is $o((\lambda - 1)^2)$ in the asymptotic regime $\lambda \downarrow 1$.

Proof. Substituting Eqs. (54) and (E2) into the functional equation (55), expanding both sides into power series in t , and comparing the coefficients of t -proportional terms, one has

$$\lambda \bar{\kappa}_1^{(m+1)} = \lambda \bar{\kappa}_1^{(m)}. \quad (\text{E5})$$

From this recurrence relation and $\bar{\kappa}_1^{(1)} = \lambda - 1$, one obtains Eq. (E3).

Similarly, comparing the coefficients of t^2 -proportional terms, one has

$$\frac{\lambda^2}{2} \bar{\kappa}_2^{(m+1)} = \frac{\lambda}{2} \bar{\kappa}_2^{(m)} + \frac{\kappa_2}{2} \{\bar{\kappa}_1^{(m)}\}^2, \quad (\text{E6})$$

which is simplified to

$$\bar{\kappa}_2^{(m+1)} = \frac{1}{\lambda} \bar{\kappa}_2^{(m)} + \frac{\kappa_2(\lambda - 1)^2}{\lambda^2}. \quad (\text{E7})$$

By solving this recurrence relation with $\bar{\kappa}_2^{(1)} = (\frac{\lambda-1}{\lambda})^2 \kappa_2$, one obtains Eq. (E4).

We prove the remaining claim via induction on k .

For $k = 3$, comparing the coefficients of t^3 -proportional terms, one has

$$\frac{\lambda^3}{6} \bar{\kappa}_3^{(m+1)} = \frac{\lambda}{6} \bar{\kappa}_3^{(m)} + \frac{\kappa_2}{2} \bar{\kappa}_1^{(m)} \bar{\kappa}_2^{(m)} + \frac{\kappa_3}{6} \{\bar{\kappa}_1^{(m)}\}^3. \quad (\text{E8})$$

From Eqs. (E3) and (E4), Eq. (E8) is simplified to

$$\bar{\kappa}_3^{(m+1)} = \frac{1}{\lambda^2} \bar{\kappa}_3^{(m)} + \frac{3\kappa_2^2 \sum_{i=1}^m \lambda^{i-1}}{\lambda^{m+4}} (\lambda - 1)^3 + \frac{\kappa_3(\lambda - 1)^3}{\lambda^3}. \quad (\text{E9})$$

By solving this recurrence relation with $\bar{\kappa}_3^{(1)} = (\frac{\lambda-1}{\lambda})^3 \kappa_3$, one obtains

$$\begin{aligned} \bar{\kappa}_3^{(m)} &= \frac{1}{\lambda^{2(m-1)}} \left(\frac{\lambda - 1}{\lambda} \right)^3 \kappa_3 \\ &\quad + \frac{(\lambda - 1) \sum_{i=1}^{2(m-1)} \lambda^{i-1}}{\lambda^{2(m-1)}} \beta_3, \end{aligned} \quad (\text{E10})$$

where

$$\begin{aligned} \beta_3 &= \frac{\lambda^2}{(\lambda - 1)(\lambda + 1)} \\ &\quad \times \left\{ \frac{3\kappa_2^2 \sum_{i=1}^m \lambda^{i-1}}{\lambda^{m+4}} (\lambda - 1)^3 + \frac{\kappa_3(\lambda - 1)^3}{\lambda^3} \right\}. \end{aligned} \quad (\text{E11})$$

From Eq. (E11) and Assumption 1, one has $\beta_3 = o(\lambda - 1)$ in the asymptotic regime $\lambda \downarrow 1$. Thus, from Eq. (E10) and Assumption 1, the third cumulant $\bar{\kappa}_3^{(m)}$ is $o((\lambda - 1)^2)$, showing that the claim holds for $k = 3$.

For $k \geq 4$, assume that, for all $l \in \{3, \dots, k - 1\}$, the l th cumulant $\bar{\kappa}_l^{(m)}$ is $o((\lambda - 1)^2)$ in the asymptotic regime $\lambda \downarrow 1$. Comparing the coefficients of t^k -proportional terms, one has

$$\frac{\lambda^k}{k!} \bar{\kappa}_k^{(m+1)} = \frac{\lambda}{k!} \bar{\kappa}_k^{(m)} + \frac{\kappa_2}{2} \sum_{l=1}^{k-1} \frac{\bar{\kappa}_l^{(m)}}{l!} \frac{\bar{\kappa}_{k-l}^{(m)}}{(k-l)!} + (**), \quad (\text{E12})$$

where the part $(**)$ is a linear combination of terms of the form $\kappa_v \prod_{j=1}^v \bar{\kappa}_{i_j}^{(m)}$, where $3 \leq v \leq k$, $\sum_{j=1}^v i_j = k$, and $1 \leq i_j \leq k - v + 1$. Simplifying Eq. (E12) yields

$$\bar{\kappa}_k^{(m+1)} = \frac{1}{\lambda^{k-1}} \bar{\kappa}_k^{(m)} + \left\{ \frac{\kappa_2}{2\lambda^k} \sum_{l=1}^{k-1} \binom{k}{l} \bar{\kappa}_l^{(m)} \bar{\kappa}_{k-l}^{(m)} + \frac{k!}{\lambda^k} (**) \right\}. \quad (\text{E13})$$

By solving this recurrence relation with $\bar{\kappa}_k^{(1)} = (\frac{\lambda-1}{\lambda})^k \kappa_k$, one obtains

$$\begin{aligned} \bar{\kappa}_k^{(m)} &= \frac{1}{\lambda^{(k-1)(m-1)}} \left(\frac{\lambda - 1}{\lambda} \right)^k \kappa_k \\ &\quad + \frac{(\lambda - 1) \sum_{i=1}^{(k-1)(m-1)} \lambda^{i-1}}{\lambda^{(k-1)(m-1)}} \beta_k, \end{aligned} \quad (\text{E14})$$

where

$$\beta_k = \frac{\lambda^{k-1}}{(\lambda-1) \sum_{i=1}^{k-1} \lambda^{i-1}} \times \left\{ \frac{\kappa_2}{2\lambda^k} \sum_{l=1}^{k-1} \binom{k}{l} \bar{\kappa}_l^{(m)} \bar{\kappa}_{k-l}^{(m)} + \frac{k!}{\lambda^k} (**) \right\}. \quad (\text{E15})$$

Regarding the first term inside the braces on the right-hand side of Eq. (E15), from Eqs. (E3) and (E4), the inductive hypothesis, and Assumption 1, its order is given as

$$\frac{\kappa_2}{2\lambda^k} \sum_{l=1}^{k-1} \binom{k}{l} \bar{\kappa}_l^{(m)} \bar{\kappa}_{k-l}^{(m)} = o((\lambda-1)^3), \quad (\text{E16})$$

which arises from the terms of the form $\kappa_2 \bar{\kappa}_1^{(m)} \bar{\kappa}_{k-1}^{(m)}$. Similarly, the second term inside the braces of Eq. (E15) is given as

$$\frac{k!}{\lambda^k} (**) = o((\lambda-1)^2), \quad (\text{E17})$$

which arises from the term of the form $\kappa_k \{\bar{\kappa}_1^{(m)}\}^k$. From Eqs. (E15)–(E17),

$$\beta_k = o(\lambda-1). \quad (\text{E18})$$

It follows from Eq. (E18) and Assumption 1(iv) that the k th cumulant $\bar{\kappa}_k^{(m)}$ is $o((\lambda-1)^2)$ in the asymptotic regime $\lambda \downarrow 1$.

Via induction, we have proved the lemma. ■

Since $\bar{\kappa}_1^{(m)} = \lambda - 1 > 0$,

$$\frac{t}{K_{W_m}(t)} = \sum_{k=0}^{\infty} \tilde{c}_k^{(m)} t^k \quad (\text{E19})$$

exists, and the coefficients $\{\tilde{c}_k^{(m)}\}$ are given by the following recurrence relation [43]:

$$\tilde{c}_0^{(m)} = \frac{1}{\bar{\kappa}_1^{(m)}}, \quad (\text{E20})$$

$$\tilde{c}_k^{(m)} = -\frac{1}{\bar{\kappa}_1^{(m)}} \left(\sum_{i=1}^k \frac{\bar{\kappa}_{i+1}^{(m)}}{(i+1)!} \tilde{c}_{k-i}^{(m)} \right). \quad (\text{E21})$$

Using Lemma E1, $\tilde{c}_0^{(m)}$ and $\tilde{c}_1^{(m)}$ are obtained as

$$\tilde{c}_0^{(m)} = \frac{1}{\lambda-1}, \quad (\text{E22})$$

$$\tilde{c}_1^{(m)} = -\frac{\kappa_2 \sum_{i=1}^m \lambda^{i-1}}{2\lambda^{m+1}}, \quad (\text{E23})$$

respectively.

Next, we prove the following lemma.

Lemma E2. Assume Assumption 1. Then, for $k \geq 2$, the coefficient $\tilde{c}_k^{(m)}$ given by Eq. (E21) with $\tilde{c}_0^{(m)} = \frac{1}{\lambda-1}$ is $o(1)$ in the asymptotic regime $\lambda \downarrow 1$.

Proof. We prove the lemma via induction on k .

For $k = 2$, from Eqs. (E21) and (E3),

$$\tilde{c}_2^{(m)} = -\frac{1}{\lambda-1} \left(\frac{\bar{\kappa}_2^{(m)}}{2!} \tilde{c}_1^{(m)} + \frac{\bar{\kappa}_3^{(m)}}{3!} \tilde{c}_0^{(m)} \right). \quad (\text{E24})$$

Thus, it follows from Eqs. (E22) and (E23), and Lemma E1 that the coefficient $\tilde{c}_2^{(m)}$ is $o(1)$ in the asymptotic regime $\lambda \downarrow 1$, showing that the claim holds for $k = 2$.

For $k \geq 3$, assume that, for all $l \in \{2, \dots, k-1\}$, the coefficient $\tilde{c}_l^{(m)}$ is $o(1)$ in the asymptotic regime $\lambda \downarrow 1$. The coefficient $\tilde{c}_k^{(m)}$ is given by Eq. (E21). Regarding the sum inside the parentheses on the right-hand side of Eq. (E21), it follows from Eqs. (E21)–(E23), the inductive hypothesis, and Lemma E1 that the sum is $o(\lambda-1)$, which arises from $\frac{\bar{\kappa}_{k+1}^{(m)}}{(k+1)!} \tilde{c}_0^{(m)}$. Therefore, by Eq. (E21), the coefficient $\tilde{c}_k^{(m)}$ is $o(1)$.

Via induction, we have proved the claim. ■

Using Eq. (E19), $h_m(t)$ defined by Eq. (58) can be written as

$$h_m(t) = \left(\sum_{k=0}^{\infty} \tilde{c}_k^{(m)} t^{k-1} \right) - \frac{1}{\lambda-1} \frac{1}{t} + \frac{\kappa_2^*}{2\lambda(\lambda-1)}. \quad (\text{E25})$$

Comparing Eq. (E25) with Eq. (62), one obtains the following correspondence between the coefficients:

$$c_0 = \tilde{c}_0^{(m)} - \frac{1}{\lambda-1} = 0, \quad (\text{E26})$$

$$\begin{aligned} c_1 &= \tilde{c}_1^{(m)} + \frac{\kappa_2^*}{2\lambda(\lambda-1)} \\ &= \frac{\kappa_2^*}{2\lambda(\lambda-1)} - \frac{\kappa_2 \sum_{i=1}^m \lambda^{i-1}}{2\lambda^{m+1}}, \end{aligned} \quad (\text{E27})$$

$$c_k = \tilde{c}_k^{(m)}, \quad k \geq 2. \quad (\text{E28})$$

Therefore, it follows from Lemma E2 that the coefficient c_k for $k \geq 2$ is $o(1)$ in the asymptotic regime $\lambda \downarrow 1$.

- [1] I. J. Bienaymé, De la loi de multiplication et de la durée des familles, *L'Institut* **13**, 131 (1845); *Soc. Philomat. Paris Extraits*, Ser. **5**, 37 (1845). (See also Sec. 1.6 of C. C. Heyde and E. Seneta, *I. J. Bienaymé: Statistical Theory Anticipated*, edited by M. J. Klein and G. J. Toomer, Studies in the History of Mathematics and Physical Sciences, Vol. 3 (Springer, 1977)).
- [2] H. W. Watson and F. Galton, On the probability of the extinction of families, *J. Anthropol. Inst. Great Britain Ireland* **4**, 138 (1875).
- [3] F. Tantillo, New class of solutions of the ultimate extinction probability from Galton–Watson processes, *Nucl. Eng. Des.* **417**, 112827 (2024).

- [4] I. A. Ventura and Q. Berger, Ising model on a Galton–Watson tree with a sparse random external field, *Electron. J. Probab.* **29**, 1 (2024).
- [5] A. Bogdanov, P. Kevei, M. Szalai, and D. Virok, Stochastic modeling of in vitro bactericidal potency, *Bull. Math. Biol.* **84**, 6 (2021).
- [6] C. J. Burden and H. Simon, Genetic drift in populations governed by a Galton–Watson branching process, *Theor. Popul. Biol.* **109**, 63 (2016).
- [7] A. G. M. Neves and C. H. C. Moreira, Applications of the Galton–Watson process to human DNA evolution and demography, *Physica A* **368**, 132 (2006).

- [8] D. E. Taneyhill, A. M. Dunn, and M. J. Hatcher, The Galton–Watson branching process as a quantitative tool in parasitology, *Parasitol. Today* **15**, 159 (1999).
- [9] F. J. Lombard and F. Martin, Statistics of electron multiplication, *Rev. Sci. Instrum.* **32**, 200 (1961).
- [10] L. A. Dietz, Basic properties of electron multiplier ion detection and pulse counting methods in mass spectrometry, *Rev. Sci. Instrum.* **36**, 1763 (1965).
- [11] J. Prescott, A statistical model for photomultiplier single-electron statistics, *Nucl. Instrum. Methods* **39**, 173 (1966).
- [12] L. A. Dietz and L. R. Hanrahan, Electron multiplier–scintillator detector for pulse counting positive or negative ions, *Rev. Sci. Instrum.* **49**, 1250 (1978).
- [13] T. E. Harris, *The Theory of Branching Processes* (Springer, Berlin, 1963).
- [14] K. B. Athreya and P. E. Ney, *Branching Processes* (Springer, Berlin, 1972).
- [15] J. L. Doob, *Stochastic Processes* (Wiley, New York, 1953).
- [16] H. Kesten and B. P. Stigum, A limit theorem for multidimensional Galton–Watson processes, *Ann. Math. Stat.* **37**, 1211 (1966).
- [17] E. Bellamy, G. Bellettini, J. Budagov, F. Cervelli, I. Chirikov-Zorin, M. Incagli, D. Lucchesi, C. Pagliarone, S. Tokar, and F. Zetti, Absolute calibration and monitoring of a spectrometric channel using a photomultiplier, *Nucl. Instrum. Methods Phys. Res. Sect. A* **339**, 468 (1994).
- [18] R. Dossi, A. Ianni, G. Ranucci, and O. Smirnov, Methods for precise photoelectron counting with photomultipliers, *Nucl. Instrum. Methods Phys. Res. Sect. A* **451**, 623 (2000).
- [19] M. Ibrahim, A. Montanari, and G. S. Moore, Accelerated time-of-flight mass spectrometry, *IEEE Trans. Signal Process.* **62**, 3784 (2014).
- [20] L. N. Kalousis, J. P. A. M. de André, E. Baussan, and M. Dracos, A fast numerical method for photomultiplier tube calibration, *J. Instrum.* **15**, P03023 (2020).
- [21] J. Weng, A. Zhang, Q. Wu, L. Ma, B. Xu, S. Qian, Z. Wang, and S. Chen, Single electron charge spectra of 8-inch high-collection-efficiency MCP-PMTs, *Nucl. Instrum. Methods Phys. Res. Sect. A* **1066**, 169626 (2024).
- [22] B. Jørgensen, Exponential dispersion models, *J. R. Statist. Soc. B* **49**, 127 (1987).
- [23] O. Barndorff-Nielsen, *Information and Exponential Families in Statistical Theory*, Wiley Series in Probability and Statistics (John Wiley & Sons, Chichester, UK, 1978).
- [24] B. Efron, *Exponential Families in Theory and Practice*, Institute of Mathematical Statistics Textbooks (Cambridge University Press, Cambridge, UK, 2023).
- [25] B. A. Sevast'yanov, *Vetvâšiesâ Processy* (Nauka, 1971).
- [26] S. Sagitov and A. Lindo, A special family of Galton–Watson processes with explosions, in *Branching Processes and Their Applications*, Lecture Notes in Statistics, edited by I. M. del Puerto, M. González, C. Gutiérrez, R. Martínez, C. Minuesa, M. Molina, M. Mota, and A. Ramos (Springer, Cham, Switzerland, 2016), Chap. 14, Vol. 219, pp. 237–254.
- [27] M. H. Holmes, *Introduction to Perturbation Methods*, Texts in Applied Mathematics, Vol. 20, 2nd ed. (Springer, New York, 2013).
- [28] E. J. Hinch, *Perturbation Methods*, Cambridge Texts in Applied Mathematics (Cambridge University Press, Cambridge, UK, 1991).
- [29] See Supplemental Material at <https://link.aps.org/supplemental/10.1103/w3yn-jpt5> for the detailed discussion of the derivation of the approximate solution $\tilde{K}(t)$ without resorting to the perturbation method.
- [30] Strictly speaking, the CPG distribution does not have a probability density function, because it has the mass $e^{-\mu} > 0$ at 0.
- [31] See Supplemental Material at <https://link.aps.org/supplemental/10.1103/w3yn-jpt5> for a proof that the CPG approximation underestimates the probability of zero of the GW model as $n \rightarrow \infty$.
- [32] H. J. Gale and J. A. B. Gibson, Methods of calculating the pulse height distribution at the output of a scintillation counter, *J. Sci. Instrum.* **43**, 224 (1966).
- [33] R. Kaas, M. Goovaerts, J. Dhaene, and M. Denuit, *Modern Actuarial Risk Theory: Using R*, 2nd ed. (Springer, Berlin, Heidelberg, Germany, 2008).
- [34] J. D. Biggins and D. N. Shanbhag, Some divisibility problems in branching processes, *Math. Proc. Cambridge Philos. Soc.* **90**, 321 (1981).
- [35] W. Feller, Diffusion processes in genetics, in *Proceedings of the 2nd Berkeley Symposium on Mathematics, Statistics, and Probability*, edited by J. Neyman (University of California Press, Berkeley and Los Angeles, California, 1951), pp. 227–246.
- [36] M. Jiřina, On Feller's branching diffusion processes, *Časopis pro pěstování matematiky* **94**, 84 (1969).
- [37] T. Lindvall, Convergence of critical Galton–Watson branching processes, *J. Appl. Probab.* **9**, 445 (1972).
- [38] T. Lindvall, Limit theorems for some functionals of certain Galton–Watson branching processes, *Adv. Appl. Probab.* **6**, 309 (1974).
- [39] R. Garcia-Millan, F. Font-Clos, and A. Corral, Finite-size scaling of survival probability in branching processes, *Phys. Rev. E* **91**, 042122 (2015).
- [40] P. Henrici, *Applied and Computational Complex Analysis Volume 1: Power Series—Integration—Conformal Mapping—Location of Zeros* (John Wiley & Sons, 1974).
- [41] K. Uemura, T. Obuchi, and T. Tanaka, Approximate modeling for supercritical Galton–Watson branching processes with compound Poisson–gamma distribution, Zenodo (2025), <https://doi.org/10.5281/zenodo.17373238>.
- [42] Here, we adopt an extended definition of the negative binomial distribution, which allows noninteger values of r .
- [43] H. S. Wilf, *Generating Functionology* (Academic Press, 1990).