


## Giant fluctuations in logistic growth of two species competing for limited resources

Bahram Houchmandzadeh

University Grenoble Alpes, CNRS, LIPhy, F-38000 Grenoble, France

 (Received 12 June 2018; revised manuscript received 23 August 2018; published 9 October 2018)

We analyze the fluctuation of the number of individuals when two species competing for the same limited resources, beginning with a few initial individuals, are submitted to a logistic growth. We show that when the total number of individuals reaches the carrying capacity, the number of each species is subject to giant fluctuations (variance  $\sim$  mean<sup>2</sup>) if the two species have similar growth rates. We show that the deterministic logistic equation can be used only when the growth rates are significantly different, otherwise such growth has to be investigated by stochastic processes tools. These results generalize to a wide class of growth law for two species competing for the same resources.

DOI: [10.1103/PhysRevE.98.042118](https://doi.org/10.1103/PhysRevE.98.042118)

### I. INTRODUCTION

In many chemical or biological systems, fluctuations can be large and drastically modify the results expected from a mean-field approximation [1]. A famous early example was investigated by Delbrück [2] for the unbounded autocatalytic chemical reaction  $A \rightarrow 2A$ , where he showed that the number  $n(t)$  of  $A$  molecules at time  $t$  displays giant fluctuations: The variance  $V(t)$  is of the order of the square of the mean  $V(t) = \langle n(t) \rangle^2 / n_0$ , where  $n_0$  is the initial number of  $A$  molecules. It can be shown that spatial diffusion is not fast enough to dilute these local fluctuations and this phenomenon can lead to spatial clustering, for example, of organisms in ecological systems [3,4] or of neutrons in nuclear reactors [5]. Other cases of large fluctuations in biological systems during unbounded growth have been considered by Das *et al.* [6].

The unbounded autocatalytic reaction captures the initial growth period but may seem unrealistic for systems where resources are limited [6]. More realistic scenarios are captured by a logistic growth where the reaction constant tends to zero as the number of replicating agents increases. If only one species is subject to such a growth, then fluctuations will become negligible when the number of replicating agents reaches the carrying capacity of the system. On the other hand, as we show below, if different species are competing for the same resources, then the number of each species can display large fluctuations similar to the above example. This situation is relevant, for example, when independent cellular pathways compete for the same resources [7], when a cell is infected initially by a few bacteria or viruses carrying different mutations, or when different mutants of cancerous cells compete with each other in the organism [8,9]. Another important example is chemical and biological reactions in small compartments such as droplets [10,11] which can be used, for example, for high-throughput directed evolution [12].

Consider the simple competition of two species of autoreplicators  $A$  and  $B$  with concentration  $x$  and  $y$ , subject to a logistic growth where their deterministic evolution equation

is given by

$$\frac{dx}{dt} = a'x(c_s - x - y), \quad (1)$$

$$\frac{dy}{dt} = b'y(c_s - x - y), \quad (2)$$

where  $a'$  and  $b'$  are their respective growth rates at small concentrations and  $c_s$  is the saturation concentration of the system. The solution of the above equations is given by

$$\frac{x}{x_0} = \left( \frac{y}{y_0} \right)^r, \quad (3)$$

where  $x_0$  and  $y_0$  are the initial concentration of each species and  $r = a'/b'$  is the relative growth rate of  $A$  in respect to  $B$  species. The final concentration of each species is found by solving  $x_\infty + y_\infty = c_s$  in combination with relation (3). In particular, for the neutral case  $r = 1$ , the final concentration of each species is explicitly given by  $x_\infty = pc_s$  and  $y_\infty = (1 - p)c_s$ , where  $p = x_0/(x_0 + y_0)$  is the initial proportion of  $A$  species.

Stochastic logistic growth have been introduced by Bartelt *et al.* [13] and studied by many authors such as Tan and Piantadosi [14] and Matis and Kiffe [15]. Consider the discrete stochastic growth process given by the following rates:

$$W(n, m \rightarrow n + 1, m) = an(N_s - N), \quad (4)$$

$$W(n, m \rightarrow n, m + 1) = bm(N_s - N), \quad (5)$$

where  $n$  and  $m$  are the number of  $A$  and  $B$  species and  $N = n + m$  is the total number of individuals at time  $t$ . The mean-field approximation of this process leads to Eqs. (1) and (2). Indeed, neglecting fluctuations by setting  $\langle n^2 \rangle = \langle n \rangle^2$ , the dynamics of the mean of each species is given by [16]

$$\frac{d\langle n \rangle}{dt} = a\langle n \rangle(N_s - \langle n \rangle - \langle m \rangle), \quad (6)$$

$$\frac{d\langle m \rangle}{dt} = b\langle m \rangle(N_s - \langle n \rangle - \langle m \rangle), \quad (7)$$

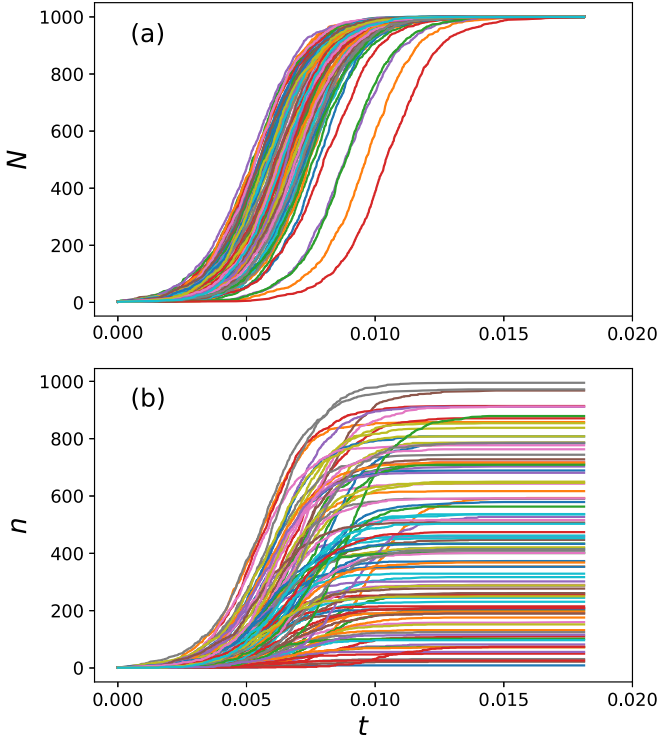


FIG. 1. Neutral logistic growth of two competing species. One hundred individual based numerical simulation of Eqs. (4) and (5) with  $a = b = 1$  and  $N_s = 1000$  are displayed. The initial number of each species is  $n_0 = m_0 = 1$ . (a) The total number of individuals  $N$  as a function of time; (b) number of individuals of species  $A$  as a function of time. Simulations are performed by a standard Gillespie algorithm (Appendix D).

setting  $x$  (respectively,  $y$ ) as  $\langle n \rangle / V$  ( $\langle m \rangle / V$ ),  $a' = aV$  ( $b' = bV$ ), and  $c_s = N_s / V$  where  $V$  is the volume of the system transforms Eqs. (6) and (7) into Eqs. (1) and (2).

Figure 1 displays the stochastic behavior of the logistic growth [Eqs. (4) and (5)] for  $r = 1$ . We observe that, as expected [6], fluctuations in the total number of individuals  $N = n + m$  disappear as  $N$  reaches the carrying capacity  $N_s$  [Fig. 1(a)]. However, the number of individuals of each species is extremely variable [Fig. 1(b)]. In fact, as we will show below, the probability of finding  $n$  individuals of type  $A$  when the system reaches saturation ( $N = N_s$ ) is *uniform* in this case  $P(n, N = N_s) = 1 / (N_s - 1)$ . For such giant fluctuations, the deterministic solution  $n_\infty = pN$  is devoid of information, and we have as much chance of finding one  $A$  individual as finding  $pN$  individuals.

In this article, we investigate analytically and numerically the stochastic Eqs. (4) and (5) in general and discuss the origin of such large fluctuations when  $r \approx 1$ . The following section is devoted to the transformation of Eqs. (4) and (5); Sec. III investigates the problem for the neutral case  $r = 1$ ; Sec. IV generalizes the solution to  $r \neq 1$ . The last section is devoted to discussion and concluding remarks. Details of some computations are given in the Appendices.

## II. MAPPING TO A SIMPLE PROBLEM

Equations (4) and (5) represent a 2+1-dimensional system where because of the nonlinearities, moment closure is lost and no closed form solution can be obtained. However, if we change the *independent* variable from time  $t$  to the total number of individuals, then the problem is mapped to a much simpler, one-dimensional one: Instead of computing the probability  $P(n, t)$  of finding  $n$  individuals of type  $A$  at time  $t$ , we compute the probability  $P(n, N)$  of finding  $n$  individuals of type  $A$  when the total number of individuals is  $N$ . For long times,  $N$  reaches the carrying capacity  $N_s$  and therefore  $P(n, t = \infty)$  and  $P(n, N = N_s)$  contain the same information. A similar transformation was recently used to compute the Luria-Delbrück distribution of the number of mutants for a general growth curve [17].

The master equation governing  $P(n, N)$  is simple (see also Appendix A). *Once* a replication event happens ( $N \rightarrow N + 1$ ), the probability that it was an  $A$  replicating ( $n \rightarrow n + 1$ ) is

$$\begin{aligned} \alpha_N^n &= \frac{W(n, m \rightarrow n + 1, m)}{W(n, m \rightarrow n + 1, m) + W(n, m \rightarrow n, m + 1)} \\ &= \frac{rn}{N + (r - 1)n}. \end{aligned}$$

The probability that it was a  $B$  replicating ( $n$  remains constant) is

$$\beta_N^n = 1 - \alpha_N^n = \frac{N - n}{N + (r - 1)n}.$$

The master equation for  $P(n, N)$  is therefore

$$P(n, N + 1) = \alpha_N^{n-1} P(n - 1, N) + (1 - \alpha_N^n) P(n, N). \quad (8)$$

At the initial time, the system contains  $N_0$  individuals,  $n_0$  of which are of type  $A$ ; the initial condition for the master equation (8) is

$$P(n, N_0) = \delta_{n_0}^n,$$

where  $\delta$  designates the Kronecker  $\delta$ . The master equation (8) is the mapping of the logistic growth into a flow problem in the  $(N, n)$  plane, where each node distributes its content  $P(n, N)$  to the adjacent ones  $(N + 1, n + 1)$  and  $(N + 1, n)$  with proportion  $\alpha_N^n$  and  $\beta_N^n$  (Fig. 2).

Because of the form of the flow, the number of  $A$  individuals  $n$  is bounded by  $n_0$  and  $N - N_0 + n_0$  (Fig. 2). Moreover, on the two boundaries, the master equation (8) reduces to a one-term recurrence relation. For example, on the lower boundary,

$$P(n_0, N + 1) = (1 - \alpha_N^{n_0}) P(n_0, N). \quad (9)$$

The probability is found to be

$$P(n_0, N) = \frac{(N_0 - n_0)_{N - N_0}}{(N_0 + sn_0)_{N - N_0}}, \quad (10)$$

where  $s = r - 1$  is the excess relative fitness of species  $A$ .  $(x)_p$  designates the Pochhammer symbol (raising factorial):

$$(x)_p = x(x + 1) \dots (x + p - 1). \quad (11)$$

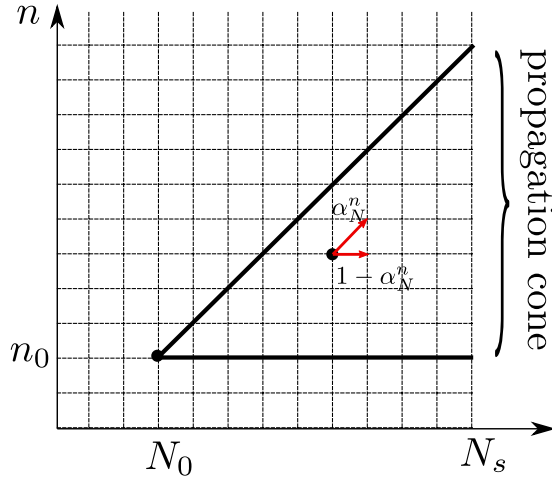


FIG. 2. Mapping of the logistic growth into a flow problem in the  $(N, n)$  plane.

Similarly, on the higher boundary,

$$P(N - N_0 + n_0, N) = \frac{(n_0)_{N-N_0}}{\left(\frac{N_0 + sn_0}{r}\right)_{N-N_0}}. \quad (12)$$

Relation (12) can also be deduced from (10) by exchanging the role of  $A$  and  $B$  individuals.

The mean of various quantities can be computed theoretically from the master equation (8). Let  $f(\cdot)$  be an arbitrary function and define

$$\langle f(n)(N) \rangle = \sum_n f(n)P(n, N),$$

then

$$\begin{aligned} \langle f(n)(N+1) \rangle &= \langle f(n)(N) \rangle + \langle \alpha_N^n [f(n+1)(N) - f(n)(N)] \rangle. \end{aligned} \quad (13)$$

For example, for  $f(n) = n$ , we have

$$\langle n(N+1) \rangle - \langle n(N) \rangle = \langle \alpha_N^n \rangle.$$

The mean-field, continuous approximation of the above expression leads to

$$\frac{d\langle n \rangle}{dN} = \alpha_N^n, \quad (14)$$

which is the equation deduced from the deterministic evolution [relations (1) and (2)].

Finally, note that it is very simple to compute numerically the probabilities obeying the master equation (8): The right-hand side of Eq. (8) is the product of a bidiagonal  $(N+1) \times N$  matrix by an  $N$ -column vector (see Appendix D).

The next two sections are devoted to the computation of the means and probabilities for the neutral and non-neutral cases.

### III. SOLUTION FOR THE NEUTRAL CASE

In the neutral case  $r = 1$ ,  $\alpha_N^n = n/N$ ; the linearity of  $\alpha$  in  $n$  allows for moment closure and exact computation of moments and probabilities. In particular, using relation (13), the mean

$\langle n(N) \rangle$  and variance  $\sigma^2(N)$  are found to obey the one-term recurrence equation

$$\langle n(N+1) \rangle = \left(1 + \frac{1}{N}\right) \langle n(N) \rangle, \quad (15)$$

$$\sigma^2(N+1) = \left(1 + \frac{2}{N}\right) \sigma^2(N) + p(1-p), \quad (16)$$

where  $p = n_0/N_0$  is the initial proportion of the  $A$  type. The two first moments are then found to be (see Appendix B 1)

$$\langle n(N) \rangle = pN, \quad (17)$$

$$\sigma^2(N) = \frac{p(1-p)}{N_0+1} N(N-N_0). \quad (18)$$

We observe that for  $N \gg N_0$ ,  $\sigma(N)$  scales *linearly* as  $N$ ; in this regime the relative amplitude of the fluctuations is

$$cv = \frac{\sigma(N)}{\langle n(N) \rangle} \approx \sqrt{\frac{1-p}{p(N_0+1)}} \quad (19)$$

and the magnitude of  $cv$  can be close to one if  $N_0$  is small. Figure 3(a) shows the perfect agreement between stochastic numerical simulations [Eqs. (4) and (5)] and the above results on the moments.

Expressions (17) and (18) can be generalized to all higher moments: Using expression (13), it can be shown (see Appendix B 1) that the raising factorial moments obey a simple relation:

$$\langle (n)_k \rangle = \langle n(n+1) \dots (n+k-1) \rangle = \frac{(n_0)_k}{(N_0)_k} (N)_k. \quad (20)$$

In the neutral case, we can go beyond moments computation and solve the master equation (8) for  $P(n, N)$ . In general,  $P(n, N|n_0, N_0)$  is a polynomial of  $n$  of degree  $N_0 - 2$ , where  $n_0, N_0$  are the initial conditions for the number of  $A$  individuals and all individuals. It is straightforward to check that (see Appendix B 2)

$$P(n, N|n_0, N_0) = A \frac{(n-n_0+1)_{n_0-1} (m-m_0+1)_{m_0-1}}{(N-N_0+1)_{N_0-1}}, \quad (21)$$

where  $m = N - n$  and, by convention,  $(x)_0 = 1$ . The normalization constant is found to be

$$A = \frac{(N_0-1)!}{(n_0-1)!(m_0-1)!}.$$

In particular,

$$P(n, N|1, 2) = \frac{1}{N-1}, \quad (22)$$

$$P(n, N|2, 3) = \frac{2(n-1)}{(N-1)(N-2)}. \quad (23)$$

The initial condition  $n_0 = 1, N_0 = 2$  was used in numerical simulations of Figs. 1 and 3.

The solution (21) is in perfect agreement with the numerical solution of the master equation (8) (Fig. 4).

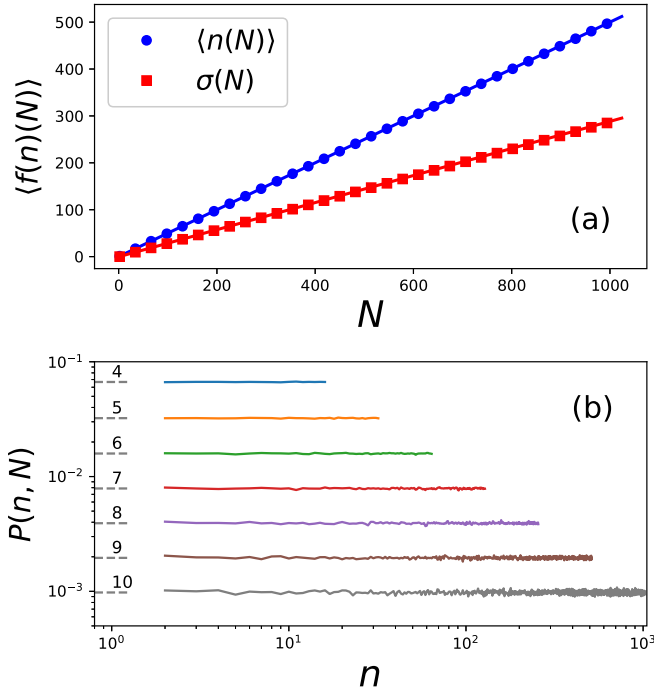


FIG. 3. Numerical stochastic simulations of equations with rates (4) and (5) (Appendix D) and comparison to theoretical values for the neutral case  $r = 1$ ,  $N_s = 1024$  and initial values  $N_0 = 2$  and  $n_0 = 1$ . (a) Evolution of the two first moments  $\langle n \rangle$  (circle) and  $\sigma$  (squares) as a function of the number of individuals  $N$ . Symbols: numerical stochastic simulations; solid lines: theoretical values given by relations (17) and (18). The moments were computed from  $M = 5000$  realizations. (b) Solid lines:  $P(n, N)$  as a function of the number of  $A$  individuals  $n$  for various values of  $N = 2^k$ ,  $k = 4, 5, \dots, 10$ . The gray dashed lines on the left designate the theoretical value  $P(n, N) = 1/(N - 1)$  (relation 22); the number above each line designates the corresponding value of  $k$ . The probabilities were computed from  $M = 10^6$  realizations.

#### IV. SOLUTION FOR $r > 1$ .

For the non-neutral case  $r > 1$ ,

$$\alpha_N^n = \frac{rn}{N + (r - 1)n}$$

is not anymore linear in  $n$  and an exact solution for  $P(n, N)$  becomes hard to obtain. However, as we are interested in the solution for large  $N$ , we can treat  $n$  and  $N$  as *continuous* variables and approximate the master equation (8) by a partial differential equation (PDE). The master equation (8) has indeed a simple structure and can be set into (see Appendix A 2)

$$\partial_N P(n, N) + \partial_n [\alpha_N^n P(n, N)] = 0. \quad (24)$$

Equation (24) is a first-order PDE and can be solved by the methods of characteristics [18]. Its general solution is found to be (see Appendix C)

$$P(n, N) = \frac{\partial}{\partial n} f \left[ \frac{(N - n)^r}{n} \right], \quad (25)$$

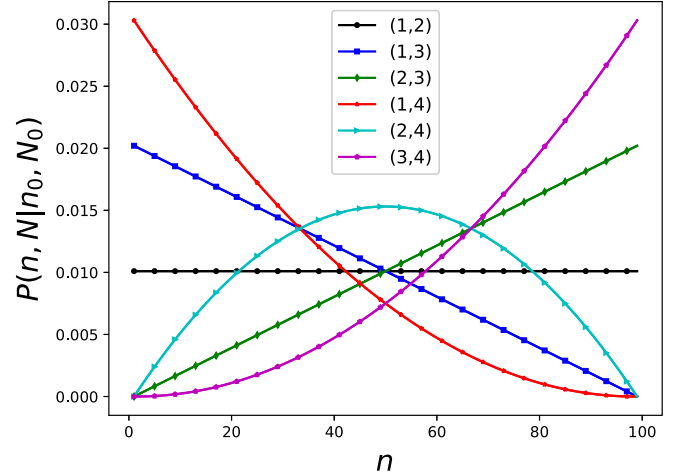


FIG. 4. The probability  $P(n, N|n_0, N_0)$  as a function of  $n$  for  $N = 100$  and various initial conditions  $(n_0, N_0)$  indicated in the legend. Solid line: theoretical solution (21); symbols: numerical solutions of the master equation (8).

where  $f(\cdot)$  is an arbitrary function to be determined from the initial condition. The implicit function  $(N - n)^r/n = \text{Cte}$  is the solution of the mean-field equation (14)  $dn/dN = \alpha_N^n$ .

Let us define  $\tilde{n}$  such that (Fig. 5)

$$\frac{(N - n)^r}{n} = \frac{(N_0 - \tilde{n})^r}{\tilde{n}}. \quad (26)$$

Then for the initial condition  $P(n, N_0) = \phi_0(n)$ , the complete solution of Eq. (24) is given by (see Appendix C)

$$P(n, N) = \frac{\partial \tilde{n}}{\partial n} \phi_0(\tilde{n}), \quad (27)$$

$$= \frac{\tilde{n}(N_0 - \tilde{n})}{N_0 + (r - 1)\tilde{n}} \frac{N + (r - 1)n}{n(N - n)} \phi_0(\tilde{n}). \quad (28)$$

No special function is defined in the mathematical literature to deal with equations of type  $x^r + ux - u = 0$ ; however, it is straightforward to find the numerical solution of Eq. (26) and use expression (28) to compute  $P(n, N)$ .

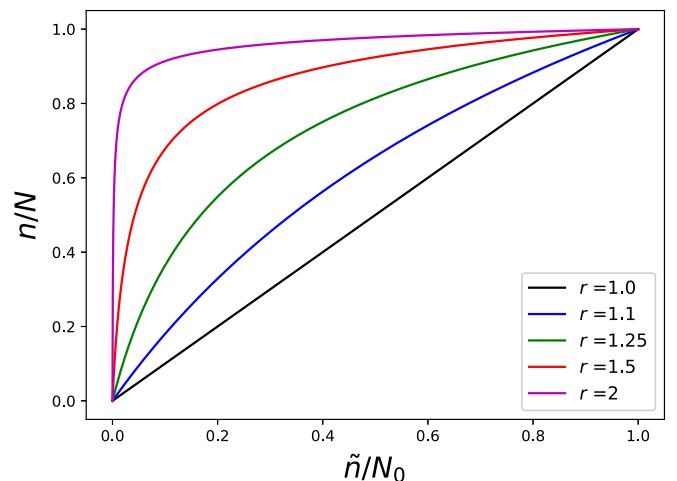


FIG. 5. Function  $n(\tilde{n})$  obtained by numerically solving the algebraic equation (26) for  $N_0/N = 10^{-3}$  and various values of  $r$ .

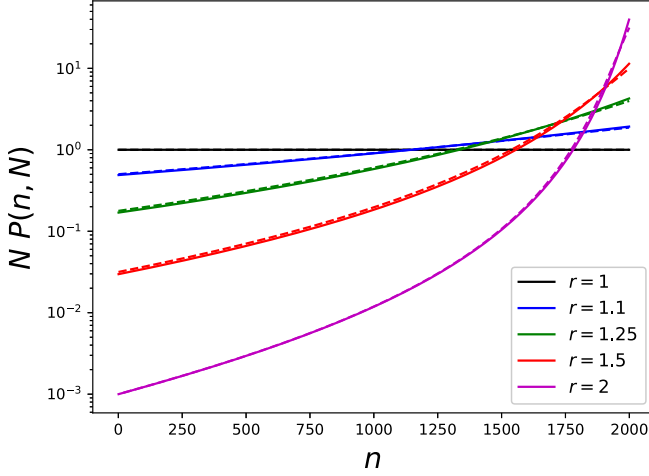


FIG. 6. Solution (32) of the continuous master equation (24) (continuous lines) compared to numerical solutions of the discrete master equation (8) (dashed lines) for  $N = 2000$ ,  $N_0 = 2$ ,  $n_0 = 1$ , and various values of  $r$ . The solution (32) is obtained by numerically solving Eq. (26) and then using relation (28).

To make it more concrete, let us consider in some details the neutral case  $r = 1$ , and compare the exact known solution (21) to the solution (27) of the PDE approach. In this case, relation (23) transforms into the explicit form  $\tilde{n} = (N_0/N)n$ . The initial condition has to be chosen in order to match the known solution (21); once it has been fixed for  $r = 1$ , it will be used for all  $r > 1$ . The initial condition corresponding to the discrete case  $n_0 = 1$ ,  $N_0 = 2$  [relation (22)] is

$$\phi_0(n) = \Pi(n - 1), \quad (29)$$

where the gate function is defined as  $\Pi(x) = 1/2$  for  $|x| < 1$  and is zero outside this domain. Therefore,

$$P(n, N) = \frac{2}{N} \Pi\left(\frac{2}{N}n - 1\right), \quad (30)$$

$$= \frac{1}{N} \quad n \in ]0, N[, \quad (31)$$

which approximates the exact solution (22) to  $O(1/N)$ .

The general solution for arbitrary  $r$  corresponding to initial condition  $n_0 = 1$ ,  $N_0 = 2$  is then simply

$$P(n, N) = \frac{1}{2} \frac{\partial \tilde{n}}{\partial n} \quad n \in ]0, N[. \quad (32)$$

Figure 6 shows the excellent agreement between expression (32) and the numerical solution obtained from the exact discrete master equation (8).

Various moments can be extracted from solution (27):

$$\langle n^k(N) \rangle_r = \int_0^N n^k P(n, N) dn = \int_0^{N_0} n^k \phi_0(\tilde{n}) d\tilde{n}, \quad (33)$$

where  $n$  inside the integrand on the right-hand side of Eq. (33) is a function of  $\tilde{n}$  through relation (26). For the neutral case  $r = 1$ ,  $n/N = \tilde{n}/N_0$  and therefore

$$\int_0^{N_0} \frac{\tilde{n}^k}{N_0^k} \phi_0(\tilde{n}) d\tilde{n} = \frac{\langle n^k(N) \rangle_1}{N^k} = \frac{(n_0)_k}{(N_0)_k} + O(1/N). \quad (34)$$

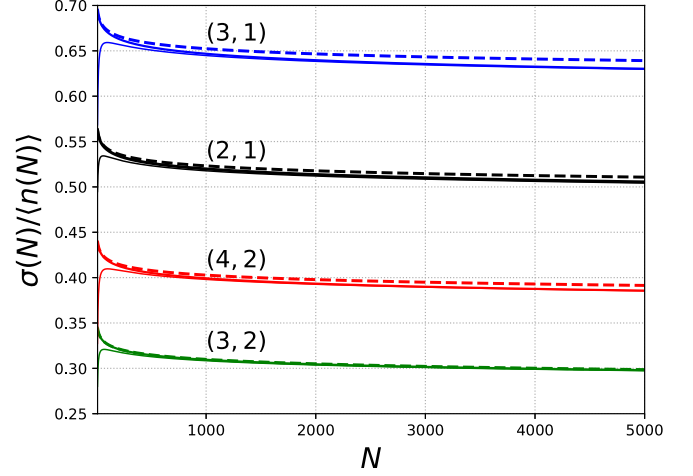


FIG. 7. Coefficient of variation  $\sigma(N)/\langle n(N) \rangle$  for  $s = 0.05$  ( $r = 1 + s$ ) and various initial conditions  $(N_0, n_0)$ . Thin solid lines: exact values obtained from direct numerical resolution of the master equation (8); Dashed lines: first-order perturbations given by expression (38); thick solid lines: second-order perturbations. The initial condition  $(N_0, n_0)$  of each curve is displayed above it.

We can obtain an explicit form of  $n$  as a function of  $\tilde{n}$  for various conditions. If  $s = r - 1 \ll 1$ , we can obtain a perturbative solution of Eq. (26) in powers of  $s$ . On the other hand, for high values of integer  $r$  such as  $r = 2, 3, 4$ , we can exactly solve the algebraic equation (26). These two cases constitute the near neutral and highly non-neutral situations and allows us to understand the general behavior of the system.

### A. Perturbative solution

Let us first consider the case  $s = r - 1 \ll 1$ . Setting  $\kappa = \log(N/N_0)$ , we have, to the second order in  $s$ :

$$x = \tilde{x} + \kappa \tilde{x}(1 - \tilde{x})s - \kappa \tilde{x}(1 - \tilde{x})[(\kappa + 1)\tilde{x} - \kappa/2]s^2, \quad (35)$$

where  $\tilde{x} = \tilde{n}/N_0$ ,  $x = n/N$ . The symmetry of Eq. (26) implies that  $\tilde{x}$  can be expressed as a function of  $x$  by simply replacing  $\kappa$  by  $-\kappa$  in expression (35). Using expressions (33) and (34) for the initial conditions  $N_0, n_0$ , to the first-order perturbations, the moments are found to be

$$\langle n(N) \rangle_r = \langle n(N) \rangle_1 \left[ 1 + \kappa s \frac{(N_0 - n_0)}{N_0 + 1} \right], \quad (36)$$

$$\sigma_r^2(N) = \sigma_1^2(N) \left[ 1 + 2\kappa s \frac{N_0 - 2n_0}{N_0 + 2} \right], \quad (37)$$

$$\frac{\sigma_r(N)}{\langle n(N) \rangle_r} = \text{cv}_1 \left[ 1 - \kappa s \frac{N_0(n_0 + 1)}{(N_0 + 1)(N_0 + 2)} \right], \quad (38)$$

where the subscript 1 refers to the neutral expressions (17)–(19). Figure 7 shows the comparison of the above expressions to exact values obtained from numerical solutions of the exact master equation (8).

We observe that the correction of the above expressions compared to neutral values Eqs. (17)–(19) are logarithmic and of the order of  $s\kappa = s \log(N/N_0)$ : The fluctuations amplitude  $\sigma$  is still large and of the order of the mean  $\langle n \rangle$ . The perturbative approach is valid for  $\kappa s \ll 1$ ; the solution for higher

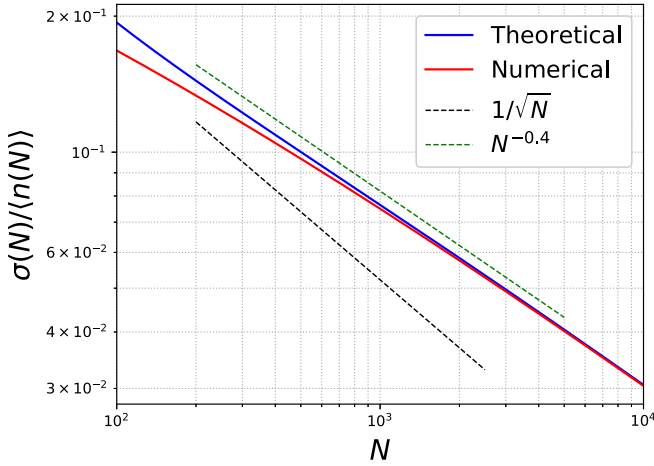


FIG. 8. Coefficient of variation  $\sigma(N)/\langle n \rangle_N$  for  $r = 2$  with initial condition  $N_0 = 2$ ,  $n_0 = 1$ . The theoretical value is obtained from expressions (41) and (42); the exact, numerical value is obtained by numerical resolution of the master equation (8). As a guide for the eye,  $N^{-0.5}$  and  $N^{-0.4}$  are also displayed.

values of  $s$  can be slightly improved by using higher-order perturbations (Fig. 7) but the perturbative approach reaches its limit for  $\kappa s \lesssim 1$ .

### B. High values of $r$

High values of  $r$  can be understood by investigating integer values such as 2, 3, and 4 for which the Eq. (26) can be exactly solved. For the case  $r = 2$

$$x = \frac{\gamma(1 - \tilde{x})^2 + 2\tilde{x} - \sqrt{\gamma^2(1 - \tilde{x})^2 + 4\gamma\tilde{x}}}{2\tilde{x}}, \quad (39)$$

where  $\gamma = N_0/N \ll 1$  and, as before,  $x = n/N$  and  $\tilde{x} = \tilde{n}/N_0$ . We will investigate the simplest case corresponding to the initial condition  $N_0 = 2$ ,  $n_0 = 1$ , where  $\phi_0(u) = \Pi(u - 1)$  [relation (29)]. For this initial condition, the moments equation (33) is greatly simplified:

$$\frac{\langle n^k \rangle}{N^k} = \frac{N_0}{2} \int_0^1 x^k d\tilde{x}. \quad (40)$$

Using expression (39), performing the integrations involved by Eq. (40) and keeping only the leading orders of  $\gamma$ , we find that

$$\begin{aligned} \frac{\langle n(N) \rangle}{N} &= 1 - \frac{4}{3}\sqrt{\gamma} + \frac{\gamma}{4}(1 - 2\log \gamma) + O(\gamma^{3/2}), \quad (41) \\ \frac{\sigma^2(N)}{N^2} &= \gamma \left( -\log \gamma - \frac{77}{18} \right) + \gamma^{3/2} \left( -\frac{4}{3}\log \gamma + \frac{106}{15} \right) \\ &\quad + O(\gamma^2). \quad (42) \end{aligned}$$

Expression (42) is valid for  $N/N_0 \gtrsim 72$ , which is indeed the regime of interest (Fig. 8). We see that for  $r = 2$ , the variance increases only as  $N \log N$  and not  $N^2$  as in the neutral case. Therefore, for high values of  $N$ , the coefficient of variation  $\sigma/\langle n \rangle$  decreases as  $(\log N/N)^{1/2}$ . In this regime, fluctuations become negligible and the deterministic approach is valid.

## V. DISCUSSION AND CONCLUSION

In this article, we have investigated the distribution of the number of individuals  $n$  and  $m$  of two species  $A, B$  during logistic growth.

The stochastic logistic growth can be considered as a specific form of competitive Lotka-Volterra growth, which has been widely discussed in the literature (see, for example, Dobrinevski and Frey [19] and references within it). Lotka-Volterra is well adapted to situations where some of the species prey on others and death rates are important. The kinds of systems we consider here, i.e., growth of pathogens inside the host cell or DNA amplification inside droplets, are devoid of these aspects and the pathogen subpopulations or DNA strands compete only through depletion of the available resources. On the relevant timescale of the growth (death of the host cell, harvesting of droplets), death phenomena are negligible and the rate equations we have considered contain only duplication events of the form  $X_i \xrightarrow{k_i} 2X_i$ , where  $k_i = a_i(N_s - N)$  denotes the exhaustion of the available resources. Finally, note that many transition rates expressions can lead to the deterministic logistic growth [(1) and (2)] but the rates (4) and (5) are the simplest ones adapted to the biological situations considered in this article.

We have shown in this article that the investigation is greatly simplified if instead of time  $t$ , the independent variable is chosen to be the total number of individuals  $N = n + m$ . This article was focused on the well-known logistic growth, but the method and conclusions are valid for any stochastic growth of the form

$$W(n, m \rightarrow n + 1, m) = anf(n, m), \quad (43)$$

$$W(n, m \rightarrow n, m + 1) = bmf(n, m), \quad (44)$$

where  $f(n, m)$  is an arbitrary function not necessarily symmetric in  $m$  and  $n$ .

The most interesting feature of the investigated system is the large amplitude of fluctuations in the neutral case  $r = a/b = 1$ , where both species have similar growth rate. Suppose that we draw (and replace)  $N_s$  individuals at random from a pool of  $N_0$  individuals when  $n_0$  are of the  $A$  type. The distribution of the number of  $A$  type in the  $N_s$  sample is a binomial one with parameter  $p = n_0/N_0$ ; the fluctuation amplitude of this experiment  $\sigma/\langle n \rangle \sim 1/\sqrt{N_s}$  is small if  $N_s \gg 1$ . One could naively suppose that a logistic growth when two types  $A$  and  $B$  individuals are competing and the system expands from  $N_0$  to  $N_s$  individuals ( $N_s \gg N_0$ ) is similar to the above drawing experiment: Each individual in the final pool draws at random its ancestor from the initial pool. This is, however, not the case and we have shown that, contrary to the binomial case, the fluctuation amplitude  $\sigma(N_s)$  scales linearly as the system size, provided  $N_s \gg N_0$ .

To get an intuitive understanding of these giant fluctuations, we note that the relative amplitude of fluctuations  $\sigma/\langle n \rangle \sim 1/\sqrt{N_0}$  can be numerically *small* if the initial population size were large and in this case, a deterministic approach captures the main behavior of the system. We can therefore subdivide the growth process into two periods: In the first period, beginning with a small population size  $(n_0, m_0)$ ,

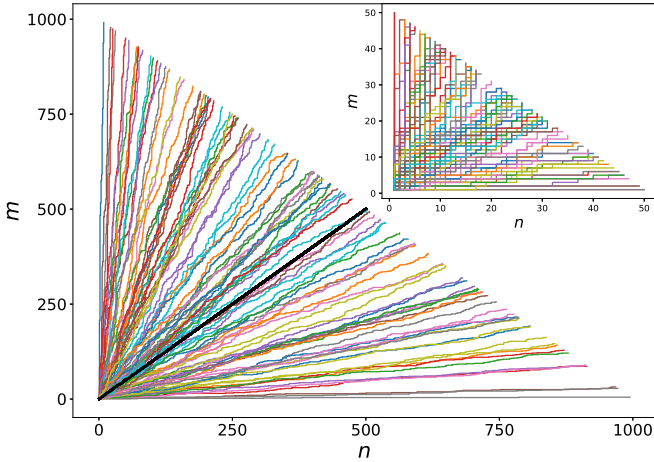


FIG. 9. Phase portrait  $(n, m)$  of individual-based simulation for the neutral case  $r = 1$ . The data are the same as in Fig. 1 but for each stochastic trajectory, at time  $t$  the number  $m$  of  $B$  individuals is reported as a function of the number  $n$  of  $A$  individuals. Thick solid black line  $m = n$  is the solution of the deterministic equations (6) and (7) with initial condition  $m_0 = n_0 = 1$ . Inset: Same data but zoomed-in to show the randomness of the initial growth.

the system is mostly stochastic and reaches a random state  $(n_1, m_1)$  where the population size cannot be considered small anymore. From this point onward, the system behaves more or less deterministically and reaches the final state corresponding to initial condition  $(n_1, m_1)$ . This intuitive understanding is best illustrated in the phase portrait of the system where  $m(t)$  is plotted against  $n(t)$  (Fig. 9) in the neutral case. We see that after an initial period when population size is small, the trajectories are approximately straight lines  $m = \kappa n$ , corresponding to the deterministic solution; on the other hand, during the initial period (Fig. 9, inset), trajectories are mostly random.

Various experiments can be devised to test the relevance of the above computations. For example, a phage such as  $\lambda$  can be modified into few different mutants, each expressing a different fluorescent proteins (such as GFP, RFP, and YFP); the mutants can then be used to co-infect a bacterial culture. The distribution of the colors in the culture after some time can be related to the probabilities we have computed through a convolution by a Poisson-Binomial distribution to account for variation in the initial number of co-infectors. A similar experiment can be performed using PCR amplification of few similar DNA strands [20] of the same length and characteristics in droplets and then analyze the number of strands copy in each droplet.

The problem we have investigated can also be used to extend the Wright-Fisher (WF) model of population genetics to variable size population (see, for example, Refs. [21–23]). In the WF model with fixed population size  $N_0$  and two mutant types  $A$  and  $B$ , each generation is formed by selecting randomly  $N_0$  individuals among the progeny of generation  $i$  to form generation  $i + 1$ . If  $x$  is the proportion of the  $A$  type with reproductive advantage  $r = 1 + s$ , then a diffusion (Kimura) equation can be derived for the evolution of the population [24,25] where the drift and diffusion coefficient are  $a(x) = sx(1 - x)$  and  $b(x) = x(1 - x)/(2N)$ .

We can generalize the WF model by allowing, at each generation  $i$ , the population to expand to size  $N_s$  and then select  $N_0$  individuals among them to form the new generation  $i + 1$ . By using the result of Sec. IV A, it is straightforward to show that the diffusion equation governing this system is the same as before except that the relative excess fitness is now renormalized to  $s' = s \log(N_s/N_0)$ . The fact that the effective fitness increases in a growing population was already noted by Ewens [21], although the amplifying factor in the problem investigated by him was proportional to the harmonic mean  $N_s$  and  $N_0$  rather than their logarithmic difference as here.

In summary, we have shown that populations subjects to logistic-like growth such as Eqs. (43) and (44) can be modeled by deterministic equations only if there is significant difference ( $r \gtrsim 2$ ) between their growth rates. If they have similar growth rate, then the deterministic equation must be abandoned and a stochastic treatment used instead.

## ACKNOWLEDGMENT

I thank Luca Peliti, David Lacoste, Marcel Vallade, Alexandre Dawid, and Hidde De Jong for fruitful discussions and critical reading of this manuscript.

## APPENDIX A: DISCRETE MASTER EQUATION DERIVATION AND ITS APPROXIMATION

### 1. Derivation

A stochastic process such as the one given by rates (4) and (5) contains essentially two different kinds of information [16]. When the system is in state  $(n, m)$ , one birth event will eventually happen and increase the total number of individuals by one unit:  $N \rightarrow N + 1$ . The timing of this event is exponentially distributed with rate

$$W = W_1 + W_2,$$

where  $W_{1,2}$  are the rates for  $n \rightarrow n + 1$  and  $m \rightarrow m + 1$  events given by (4) and (5). On the other hand, once one birth occurs, the probability that it was a type  $i$  that duplicated is

$$p_i = W_i / W.$$

In particular, the probability that it was an  $A$  (respectively,  $B$ ) individual duplicating is (conditioned on the occurrence of one birth)

$$p(n \rightarrow n + 1 | N \rightarrow N + 1) = \alpha_N^n = \frac{rn}{N + (r - 1)n},$$

$$p(m \rightarrow m + 1 | N \rightarrow N + 1) = \beta_N^n = 1 - \alpha_N^n.$$

The probability  $P(n, N + 1)$  of having  $n$  individuals of type  $A$  (and  $m = N + 1 - n$  of type  $B$ ) when there are  $N + 1$  individuals present is as follows: (i) either there were  $n - 1$  individuals when the system size was  $N$  AND when a duplication occurred it was an  $n - 1 \rightarrow n$  event OR (ii) there were  $n$  individuals when the system size was  $N$  AND when a duplication occurred it was a  $B$  duplicating event ( $n$  remains constants). Therefore,

$$P(n, N + 1) = P(n - 1, N)\alpha_N^{n-1} + P(n, N)\beta_N^n, \quad (\text{A1})$$

which is the discrete master equation (8).

## 2. Approximation by a PDE

When  $r > 1$ , the discrete master equation cannot be solved exactly anymore, but it can be approximated by a PDE in a Van Kampen's like expansion. We assume that  $N_s \gg 1$  and use it as the system size. Let us define  $\ell = 1/N_s$ ,  $x = n/N_s$ , and  $z = N/N_s$ . Furthermore, we set

$$P(n, N) = \ell^2 p(x, z). \quad (\text{A2})$$

Note that the expression of  $\alpha$  is conserved in the new variables  $x, z$ :

$$\alpha_N^n = \frac{rn}{N + (r-1)n} = \frac{rx}{z + (r-1)x} = \alpha(x, z).$$

In the new variables  $x, z$ , the master equation (A1) becomes

$$\begin{aligned} & \ell^2 [p(x, z + \ell) - p(x, z)] \\ &= \ell^2 [\alpha(x - \ell, z)p(x - \ell, z) - \alpha(x, z)p(x, z)] \end{aligned}$$

expanding this expression to the leading order in  $\ell$  leads to

$$\frac{\partial p}{\partial z} = -\frac{\partial(\alpha p)}{\partial x}, \quad (\text{A3})$$

which is a PDE in "continuous" variable  $x, z$ . We can solve this equation by the technique detailed in Appendix C to obtain  $p(x, z)$  and then get back to  $P(n, N)$  through relation (A2). On the other hand, we can directly multiply both side of Eq. (A3) by the factor  $\ell^2/\ell = 1/N_s$  and write it directly as

$$\partial_N P(n, N) = -\partial_n [\alpha_N^n P(n, N)],$$

where  $n$  and  $N$  are considered as continuous. The two approach are strictly similar but the latter allows more direct comparison to the known results and has been preferred in this article.

## APPENDIX B: VARIOUS NEUTRAL COMPUTATIONS

### 1. Factorial moments

Consider the function  $f(n) = (n)_k = n(n+1)\dots(n+k-1)$ ; then

$$f(n+1) - f(n) = (n+1)_{k-1}(n+k-n) = k(n+1)_{k-1}$$

and therefore

$$n[f(n+1) - f(n)] = k(n)_k = kf(n).$$

Therefore, using the general expression (13), we find the one term recurrence relation

$$\langle f(n)(N+1) \rangle = \left(1 + \frac{k}{N}\right) \langle f(n)(N) \rangle. \quad (\text{B1})$$

The solution of one term recurrence relations such as

$$y_{N+1} = a_N y_N$$

is

$$y_N = \left( \prod_{\ell=N_0}^{N-1} a_\ell \right) y_{N_0}. \quad (\text{B2})$$

For example, for the mean, i.e., the first factorial moment with  $k = 1$ ,

$$a_N = 1 + \frac{1}{N} = \frac{N+1}{N} \quad (\text{B3})$$

the parenthesis term in relation (B2) is of the form

$$\frac{N_0+1}{N_0} \times \frac{N_0+2}{N_0+1} \times \dots \times \frac{N}{N-1} = \frac{N}{N_0}$$

and therefore the mean is

$$\langle n(N) \rangle = (N/N_0)n_0 = pN,$$

which is the expression (15). This computation is easily generalized to higher factorial moments ( $k > 1$ ) and leads to expression (20). The variance can be deduced from the expression

$$\sigma^2 = \langle (n)_2 \rangle - \langle n \rangle^2 - \langle n \rangle$$

and leads to expression (18).

### 2. Expression of the probability

To shorten the notations, we use  $m = N - n$  whenever needed. The master equation in the neutral case is

$$P(n, N+1) = \frac{n-1}{N} P(n-1, N) + \frac{m}{N} P(n, N). \quad (\text{B4})$$

Consider

$$P(n, N) = \frac{(n-n_0+1)_{n_0-1} (m-m_0+1)_{m_0-1}}{(N-N_0+1)_{N_0-1}}. \quad (\text{B5})$$

Pochhammer manipulation is similar to factorial manipulation. In particular,

$$N(N-N_0+1)_{N_0-1} = (N-N_0+1)_{N_0}$$

$$(n-1)(n-n_0)_{n_0-1} = (n-n_0)_{n_0}$$

$$m(m-m_0+1)_{m_0-1} = (m-m_0+1)_{m_0}$$

and therefore, the right-hand side of relation (B4) is found to be

$$\frac{(n-n_0+1)_{n_0-1} (m-m_0+2)_{m_0-1}}{(N-N_0+1)_{N_0}} (n-n_0+m-m_0+1).$$

As

$$n-n_0+m-m_0+1 = N-N_0+1$$

and

$$\frac{N-N_0+1}{(N-N_0+1)_{N_0}} = \frac{1}{(N+1-N_0+1)_{N_0-1}},$$

expression (B5) is indeed a solution of the master equation, up to a multiplicative constant. The constant is found by stating  $P(n_0, N_0) = 1$ . As the master equation conserves the probability, the constant is valid for all  $N$ .

### APPENDIX C: SOLVING THE PDE

Consider a first-order PDE of first order for the function  $P(x, t)$  of type

$$\partial_t P + \partial_x(\alpha P) = 0, \quad (\text{C1})$$

where  $\alpha = \alpha(x, t)$  is a known function. Let us call  $R(x, t) = \text{Cte}$  the solution of the characteristic equation

$$\frac{dx}{dt} = \alpha(x, t).$$



Then, by definition,

$$\partial_t R + \alpha \partial_x R = 0.$$

Consider the function

$$Q(x, t) = \frac{\partial}{\partial x} f(R(x, t)), \quad (\text{C2})$$

where  $f(\cdot)$  is an arbitrary function. Then

$$\partial_t Q + \partial_x(\alpha Q) = \partial_x\{(\partial_t R + \alpha \partial_x R)f'(R)\} = 0$$

and therefore  $Q(x, t)$  is a solution of Eq. (C1). For example, for  $\alpha = c$ , the solution is the trivial propagation  $P(x, t) = f(x - ct)$ .

The function  $f(\cdot)$  has to be determined from the initial condition  $P(x, t_0) = \phi_0(x)$ . Consider two points  $(t_0, \tilde{x})$  and  $(t, x)$  in the plane, related through  $R(x, t) = R(\tilde{x}, t_0)$ , i.e., they belong to the same characteristic curve. Obviously, we can reverse this relation as  $\tilde{x} = g(R(x, t), t_0)$  and therefore write the general solution (C2) as  $P(x, t) = \partial_x f(\tilde{x}) = (\partial \tilde{x} / \partial x) f'(\tilde{x})$ . On the other hand, at the initial time  $t_0$ ,  $x = \tilde{x}$ ,  $\partial \tilde{x} / \partial x = 1$ , and therefore  $f'(\cdot) = \phi_0(\cdot)$ . The solution of the PDE (C2) with the initial condition  $\phi_0(x)$  is then

$$P(x, t) = \frac{\partial \tilde{x}}{\partial x} \phi_0(\tilde{x}).$$

$P(\cdot, t)$  can be seen as a transformation, i.e., scaling and deformation of the initial condition  $\phi_0(\cdot)$ . An initial Dirac distribution, however, propagates without deformation along a characteristic curve because  $f(x)\delta(x) = \delta(x)$ : In this case, the PDE is reduced to the deterministic equations  $dx/dt = \alpha$ .

Let us further define the function  $\phi_0(\cdot)$  used in this article for the PDE (24). The true probability  $P_d(n, N)$  is function of *discrete* variables  $n$  and  $N$ . In order to estimate this probability, we have used the probability density  $P_c(n, N)$  of *continuous* variable  $n, N$ .  $P_c$  must approximate  $P_d$  for large  $N$ .

$\phi_0(\cdot)$  has to be chosen to make this approximation as precise as possible. However, we cannot use the discrete initial condition  $P(n, N_0) = \delta_{n_0}^n$ , because the continuous PDE will be reduced to a deterministic equation. We make the assumption that the choice of  $\phi_0(\cdot)$  is independent of  $r$  and therefore can be deduced from the known expression of neutral probability. For  $r = 1$ ,  $\tilde{n} = (N_0/N)n$ , and therefore we have

$$\phi_0(\tilde{n}) = \frac{N}{N_0} P_1\left(\frac{N}{N_0} \tilde{n}, N\right),$$

where  $P_1(\cdot)$  is the neutral probabilities but the arguments are continuous.

#### APPENDIX D: NUMERICAL SIMULATIONS

The individual-based numerical simulations of Eqs. (4) and (5) (such as shown in Fig. 1) are performed using a Gillespie algorithm [26] written in the C language. In short, when the system is in state  $(m, n)$ , the time to the next duplication is drawn from an exponential distribution with rate  $S = W_1 + W_2$ , where  $W_{1,2}$  are the rates (4) and (5) and the total population size  $N = n + m$  is incremented. The decision that this duplication is an  $n \rightarrow n + 1$  event is drawn from a [0,1] uniform probability with weight  $W_1/S$ . A trajectory is computed by iterating this process, beginning with initial condition  $(n_0, m_0)$ .  $M$  different trajectories are then generated to perform the necessary statistics.

The numerical solution of the master equation (8) is performed by a simple iteration written in the high-level language Julia [27]. The basic structure is an  $N_s \times N_s$  matrix  $P$  where the element  $P_{n,N}$  represent  $P(n, N)$ . The solution is computed by iterations, where column  $N + 1$  is deduced from column  $N$  according to Eq. (8) and illustrated in Fig. 2.

- 
- [1] L. S. Tsimring, Noise in biology, *Rep. Prog. Phys.* **77**, 026601 (2014).
- [2] M. Delbrück, Statistical fluctuations in autocatalytic reactions, *J. Chem. Phys.* **8**, 120 (1940).
- [3] B. Houchmandzadeh, Neutral Clustering in a Simple Experimental Ecological Community, *Phys. Rev. Lett.* **101**, 078103 (2008).
- [4] B. Houchmandzadeh, Theory of neutral clustering for growing populations, *Phys. Rev. E* **80**, 051920 (2009).
- [5] E. Dumonteil, F. Malvagi, A. Zoia, A. Mazzolo, D. Artusio, C. Dieudonné, and C. De Mulatier, Particle clustering in Monte Carlo criticality simulations, *Ann. Nucl. Energy* **63**, 612 (2014).
- [6] D. Das, D. Das, and A. Prasad, Giant number fluctuations in microbial ecologies, *J. Theor. Biol.* **308**, 96 (2012).
- [7] A. J. Genot, T. Fujii, and Y. Rondelez, Computing with Competition in Biochemical Networks, *Phys. Rev. Lett.* **109**, 208102 (2012).
- [8] J. A. Spratt, D. von Fournier, J. S. Spratt, and E. E. Weber, Decelerating growth and human breast cancer, *Cancer* **71**, 2013.
- [9] N. C. Atuegwu, L. R. Arlinghaus, X. Li, A. B. Chakravarthy, V. G. Abramson, M. E. Sanders, and T. E. Yankeelov, Parameterizing the logistic model of tumor growth by DW-MRI and DCE-MRI data to predict treatment response and changes in breast cancer cellularity during neoadjuvant chemotherapy, *Transl. Oncol.* **6**, 256 (2013).
- [10] H. Song, D. L. Chen, and R. F. Ismagilov, Reactions in Droplets in Microfluidic Channels, *Angew. Chem., Int. Ed.* **45**, 7336 (2006).
- [11] A. Baccouche, S. Okumura, R. Sieskind, E. Henry, N. Aubert-Kato, N. Bredeche, J.-F. Bartolo, V. Taly, Y. Rondelez, T. Fujii, and A. J. Genot, Massively parallel and multiparameter titration of biochemical assays with droplet microfluidics, *Nat. Protocols* **12**, 1912 (2017).
- [12] J. J. Agresti, E. Antipov, A. R. Abate, K. Ahn, A. C. Rowat, J.-C. Baret, M. Marquez, A. M. Klibanov, A. D. Griffiths, and D. A. Weitz, Ultrahigh-throughput screening in drop-based microfluidics for directed evolution, *Proc. Natl. Acad. Sci. USA* **107**, 4004 (2010).
- [13] M. S. Bartlett, J. C. Gower, and P. H. Leslie, A comparison of theoretical and empirical results for some stochastic population models, *Biometrika* **47**, 1 (1960).
- [14] W. Y. Tan and S. Piantadosi, On stochastic growth processes with application to stochastic logistic growth, *Stat. Sin.* **1**, 527 (1991).

- [15] J. H. Matis and T. R. Kiffe, On approximating the moments of the equilibrium distribution of a stochastic logistic model, *Biometrics* **52**, 980 (1996).
- [16] C. Gardiner, *Handbook of Stochastic Methods: For Physics, Chemistry and the Natural Sciences* (Springer, Berlin, 2004).
- [17] B. Houchmandzadeh, General formulation of Luria-Delbrück distribution of the number of mutants, *Phys. Rev. E* **92**, 012719, (2015).
- [18] A. D. Polyanin, V. F. Zaitsev, and A. Moussiaux, *Handbook of First-Order Partial Differential Equations* (CRC Press, New York, 2001).
- [19] A. Dobrinevski and E. Frey, Extinction in neutrally stable stochastic Lotka-Volterra models, *Phys. Rev. E* **85**, 051903 (2012).
- [20] Y. Schaerli, R. C. Wootton, T. Robinson, V. Stein, C. Dunsby, M. A. A. Neil, P. M. W. French, A. J. deMello, C. Abell, and F. Hollfelder, Continuous-flow polymerase chain reaction of single-c DNA in microfluidic microdroplets, *Anal. Chem.* **81**, 302 (2009).
- [21] W. J. Ewens, The probability of survival of a new mutant in a fluctuating environment, *Heredity* **22**, 438 (1967).
- [22] S. P. Otto and M. C. Whitlock, The probability of fixation in populations of changing size, *Genetics* **146**, 723 (1997).
- [23] K. Wienand, E. Frey, and M. Mobilia, Evolution of a Fluctuating Population in a Randomly Switching Environment, *Phys. Rev. Lett.* **119**, 158301 (2017).
- [24] M. Kimura, Solution of a process of random genetic drift with a continuous model, *Proc. Natl. Acad. Sci. USA* **41**, 144 (1955).
- [25] W. J. Ewens, *Mathematical Population Genetics* (Springer-Verlag, Berlin, 2004).
- [26] D. T. Gillespie, Exact stochastic simulation of coupled chemical reactions, *J. Phys. Chem.* **81**, 2340 (1977).
- [27] J. Bezanson, A. Edelman, S. Karpinski, and V. B. Shah, Julia: A Fresh Approach to Numerical Computing, *SIAM Rev.* **59**, 65 (2017).