

**Markov-chain approach to the distribution of ancestors in species of biparental reproduction**

M. Caruso\*

*Departamento de Física Teórica y del Cosmos, Universidad de Granada, Campus de Fuentenueva, Granada (18071), España*

C. Jarne†

*Departamento de Física, Facultad de Ciencias Exactas, IFLP-CONICET, Universidad Nacional de La Plata, La Plata (1900), C.C.67, Argentina*

(Received 21 October 2013; revised manuscript received 23 April 2014; published 21 August 2014)

We studied how to obtain a distribution for the number of ancestors in species of sexual reproduction. Present models concentrate on the estimation of distributions repetitions of ancestors in genealogical trees. It has been shown that it is not possible to reconstruct the genealogical history of each species along all its generations by means of a geometric progression. This analysis demonstrates that it is possible to rebuild the tree of progenitors by modeling the problem with a Markov chain. For each generation, the maximum number of possible ancestors is different. This presents huge problems for the resolution. We found a solution through a dilation of the sample space, although the distribution defined there takes smaller values with respect to the initial problem. In order to correct the distribution for each generation, we introduced the invariance under a gauge (local) group of dilations. These ideas can be used to study the interaction of several processes and provide a new approach on the problem of the common ancestor. In the same direction, this model also provides some elements that can be used to improve models of animal reproduction.

DOI: [10.1103/PhysRevE.90.022125](https://doi.org/10.1103/PhysRevE.90.022125)

PACS number(s): 02.50.Fz, 87.23.Kg, 87.10.Ca, 89.75.Hc

**I. INTRODUCTION**

Until now, previous attempts aiming to calculate the number of ancestors in species of sexual reproduction have not been totally successful. Present models concentrate on the estimation of distributions of ancestors repetitions in genealogical trees [1–3]. It has been shown that it is not possible to reconstruct the genealogical history of each species along all its generations by means of a geometric progression [4]. The reason is that the geometric progression is determined by a sequence of independent events. We postulate that *blood relationship* is a kind of interaction that connects the events. It is possible to rebuild the tree of progenitors by modeling the problem with a Markov chain. If we consider a random variable that represents the number of ancestors present in a given generation, the size of the sample space depends on each generation. This brings serious complications on the solution of the problem, not only of a mathematical nature. We propose submerging the original sample space into a larger one. This *dilution* modifies the probability distribution. We show the need to implement a covariant derivative, due to a gauge transformation, which leaves the evolution equation invariant and corrects the probability distribution.

The main goal of the present work is to describe the distribution of ancestors for species with sexual reproduction, but also to show the novel method used here to solve other stochastic problems.

There are two important assumptions about the biology of the considered system. The first one is that the species described here has no specific behavior of sexual partner selection (random mating reproduction) [5,6]. Many species or population groups exhibit this kind of reproduction. This is

the most simple case to perform the calculation. The second assumption is about the population size. The distribution of ancestors for a given generation is contained in a population large enough to not force the selection of blood-related sexual partners. The partners could be blood related or not, randomly. Current model presents a random mating in nonoverlapping generations with negligible mutation and selection. These two assumptions are common to develop population genetic models; in particular, these are present in the Hardy-Weinberg principle [7,8].

In this work we show a way to calculate a probability distribution to get a certain number of ancestors for each generation. We have obtained its first two cumulants: the expected value of the number and its dispersion.

Specific conditions about small-size populations or specific sexual behavior can be considered later as modifications of the general case described here.

The present work will be useful in order to understand the origin of species extrapolating the individual genealogy for all members at the beginning of the species. It is possible to go one step further, to establish how populations can be affected by certain conditions, such as isolation or migration of individuals, by studying population groups with different genetic pools [9]. These ideas can be used to perform more realistic models in animal populations and also to improve estimations about extinction processes.

**II. A MARKOVIAN APPROACH TO THE ANCESTORS PROBLEM**

To calculate the number of ancestors of an individual it is necessary to use a statistical approach. If we simply accept that  $2^{t+1}$  allows us to calculate the number of ancestors in the  $t$  generation, where  $t = 0$  is the generation of progenitors of the first order (or parents for short) and so forth, we arrive at an absurdity. As we turn to past generations, the probability that

\*mcaruso@ugr.es

†jarne@fisica.unlp.edu.ar

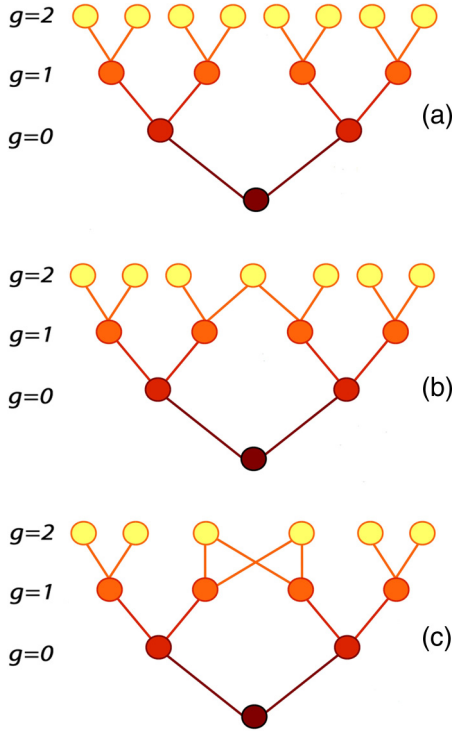


FIG. 1. (Color online) Examples of three kinds of genealogical trees (only the first few  $g$  generations). (a) No restrictions by blood relationship. (b, c) Two kinds of restriction in third generation: ancestors sharing one parent (b) or ancestors sharing two parents (c). The restriction by blood relationship increases according to the degree of endogamy.

some ancestors have been relatives is significantly larger [1–3]. This implies a restriction on the number of ancestors with respect to  $2^{t+1}$ . This last quantity corresponds to the maximum possible number of ancestors in each  $t$  generation.

There are several examples showing different ways in which the number of ancestors is reduced with respect to the maximum number in each generation. As stated in Ref. [4], the reduction of the number of ancestors, compared to  $2^{t+1}$ , is caused by blood relationship. Figure 1 shows, as an example, only the three first generations of a genealogical tree with two different ways to constrain the number of ancestors. There is a way to weigh the blood relationship using a statistical approach that includes all possible kinds of relationship in each generation. In this approach the only constraint in the number of ancestors is caused by random blood relationship between individuals of the same generation. We considered a population of ancestors whose maximum size in each  $t$  generation is given by the geometric progression  $2^{t+1}$ .

We did not consider any restriction for the number of ancestors generated by issues related to culture, in the human case, ethological in the animal case, or isolation of populations, etc. If we want to study the distributions of ancestors of individuals from populations where there are less or equal individuals than  $2^{t+1}$  for the  $t$  generation, there is an additional restriction on the number of ancestors. Blood relationship interconnects the events in the original process that leads

to  $2^{t+1}$ , which was generated by independent events and no relation between ancestors of each generation.

Derrida’s model [1,2] is based on numerical simulation under the same assumptions (closed population evolving under sexual reproduction with nonoverlapping generations). The population size is fixed for all generations and equally divided into two groups, representing males and females. At every generation, they form random heterosexual pairs and assign them a certain number of descendants according to a Poisson distribution. This is done by choosing for each male or female one pair of parents at random in the previous generation.

In our work the population size is not fixed, but always bigger than  $2^{t+1}$  for each  $t$  generation.

We defined two random variables  $y(t)$  and  $x(t)$ , which represents the number of individuals who are inside and outside to the set of ancestors, with respect to the maximum possible number of ancestors in each  $t$  generation. For this definition we have

$$x(t) + y(t) = 2^{t+1}. \tag{1}$$

We considered each generation as a link of the chain that forms a first-order Markov process [10,11]. This process is constructed on a given set of individuals ordered by generations. We take the current generation and we count its parents. Then we take all selected individuals and remake the previous question and so forth. There exists a generation in which the question or previous classification makes no more sense, in which case the process ends after a given generation. This kind of process is widely used to describe the evolution of traits that adopt a finite number of states [12].

From Eq. (1) we have

$$y(t) = 2^{t+1} - x(t), \tag{2}$$

if  $y(t)$  describes a Markov process, then that implies  $x(t)$  describes another Markov process.

We do not distinguish the different kinds of blood relationship between the ancestors of a particular generation such as brothers or cousins, and so forth. We simply consider them as indistinguishable and we just count how many there are. For the purpose of the calculations we consider  $t$  as a continuous variable. Finally, we associate a discrete-time Markov process to the continuous-time Markov process  $\{x(t) : t \geq 0\}$  called a *skeleton process* [13] defined as  $\{x(g) : g \geq 0\}$ , where  $g$  is the generation number.

The time evolution of this process is determined by the knowledge of the probability distribution in each  $t$  generation, denoted by

$$p_n(t) = \mathbb{P}[x(t) = n], \tag{3}$$

for all  $(n, t) \in \mathbb{S}_t \times \mathbb{R}$ , where  $\mathbb{S}_t$  is the sample space of  $x(t)$ , which corresponds to the interval  $[0, n_t]$ ,  $n_t = \lfloor 2^{t+1} \rfloor - 2$  and  $\lfloor z \rfloor$  is the integer part of a real number  $z$ .

An equivalent way to describe the process is through an initial value  $p_n(0)$  and the conditional probability given by  $\mathcal{P}_{nm}(t, s) = \mathbb{P}[x(t) = n | x(s) = m]$ , which represents the transition matrix elements of the states  $(m, s) \mapsto (n, t)$ .

For each generation the events are mutually exclusive. Consequently, at the time  $t + \epsilon$  the probability of finding  $n$  restrictions is given by the transition from  $m$  restrictions at

time  $t$ ; in this way,

$$p_n(t + \epsilon) = \sum_{m \in \mathbb{S}_t} \mathcal{P}_{nm}(t + \epsilon, t) p_m(t). \quad (4)$$

After some elementary operations (see Appendix A) we get

$$d_t p_n(t) = \sum_{m \in \mathbb{S}_t} \mathcal{Q}_{nm}(t) p_m(t), \quad (5)$$

where  $d_t$  denotes the total time derivative  $\frac{d}{dt}$ ,  $\mathcal{Q}_{nm}(t) = \partial_t \mathcal{P}_{nm}(t, s)|_{s=t}$  is called the *infinitesimal generator* and  $\delta_{nm}$  is the Kronecker delta.

We define  $\boldsymbol{\varphi}(t)$  as an  $|\mathbb{S}_t|$ -tuple of the probability distribution as  $\boldsymbol{\varphi}(t) = [p_0(t), p_1(t), \dots, p_{n_t}(t)]^\top$ , where  $|\mathbb{S}_t|$  denotes the cardinal number of  $\mathbb{S}_t$  and  $^\top$  represents the transposition.

The evolution equation for the process can be expressed in a *matrix form* as

$$d_t \boldsymbol{\varphi}(t) = \mathbf{Q}(t) \boldsymbol{\varphi}(t). \quad (6)$$

We denote the expectation number of ancestors by  $\alpha(t) = \langle y(t) \rangle$  and from Eq. (2),

$$\alpha(t) = 2^{t+1} - \langle x(t) \rangle, \quad (7)$$

where  $\langle x^k(t) \rangle$  is the expectation value of  $x(t)$  raised to the positive integer power  $k$  (or  $k$  moment for short) of the distribution  $p_n(t)$  and by definition is  $\langle x^k(t) \rangle = \sum_n n^k p_n(t)$ . The quantity  $\langle x(t) \rangle$  represents a constraint caused by blood relationship, which affects the expectation number of ancestors in each generation.

### III. DILUTION OF SAMPLE SPACE VIA GAUGE GROUP OF DILATIONS

The sample space of  $x(t)$  is different for each  $t$  generation, thus there is enormous difficulty to solve Eq. (6). We considered a *dilution* of  $\mathbb{S}_t$  within a larger set  $\mathbb{S} \supseteq \mathbb{S}_t$ , for all  $t$ , consisting of replacing the endpoint  $n_t$  by a huge number  $N$ . This dilution can be viewed as a *dilation* represented with the substitution rule  $n_t \mapsto N$ , such that  $\mathbb{S} = [0, N]$ . On the other hand, we know that there exists a certain  $T$  generation that can be considered as the end of the process. The existence of a limit generation,  $T$ , allows us to choose  $N = n_T$ . Consequently, we can solve the problem in this *dilated sample space* and then recover the lost endpoint caused by the dilation through a suitable transformation. The price to pay for it is the need of renormalization of the distribution defined on  $\mathbb{S}$  to compensate for the dilution effect. The renormalization takes place by a linear transformation, which modifies the norm of the distribution for each generation. This *local* transformation (i.e., depends on each  $t$ ), is structured as a *gauge group*, specifically the *group of local dilations*. Essentially, the distribution defined on  $\mathbb{S}_t$  is equivalent to the renormalized distribution, which is defined on the dilated sample space  $\mathbb{S}$ .

In summary, we can interpret that the process on  $\mathbb{S}_t$  is the result of a process on this larger set  $\mathbb{S}$ , which interacts with another process on the complement of  $\mathbb{S}_t$ , denoted by  $\mathbb{S} - \mathbb{S}_t$ . This interaction is represented by the renormalization of the distribution defined on  $\mathbb{S}$ , in an effective theory context. As long as the process on  $\mathbb{S}$  becomes much simpler, the description

on  $\mathbb{S} - \mathbb{S}_t$  will be more complex. This is the basis for the dilation transformation, which is discussed in Appendices E and F.

We considered a version in which the sample space  $\mathbb{S}_t$  is dilated to the set of natural numbers  $\mathbb{N}$ , including the 0 element. Then we have only one boundary condition for the state  $n = 0$ . This allows us to focus on *time homogeneous processes*; i.e., the infinitesimal generator is independent of  $t$ . Another consideration is the *spatial homogeneity*, i.e., the case where the infinitesimal generator does not depend on the state of the random variable  $X(t)$ .

The Markov process in this larger sample space  $\mathbb{N}$  requires us to consider two new random variables  $\{X, Y\}$ , defined on  $\mathbb{N}$  and related in a similar way to the old random variables  $\{x, y\}$  from Eq. (2). The associated probability distribution is denoted by  $P_n(t) = \mathbb{P}[X(t) = n]$  and defines  $\boldsymbol{\phi}(t) = [P_0(t), P_1(t), \dots]^\top$ , which satisfies the equation

$$d_t \boldsymbol{\phi}(t) = \mathbf{Q} \boldsymbol{\phi}(t). \quad (8)$$

Knowing the initial conditions  $\boldsymbol{\phi}(0) = (1, 0, \dots)^\top$  and the infinitesimal generator  $\mathbf{Q}$ , we can write the formal solution of Eq. (8) as

$$\boldsymbol{\phi}(t) = \exp(t\mathbf{Q}) \boldsymbol{\phi}(0). \quad (9)$$

In order to establish the matrix  $\mathbf{Q}$ , we study the time evolution  $t \mapsto t + \epsilon$  for small value of  $\epsilon$ . Therefore, only transitions to the nearest states are allowed, because the infinitesimal time evolution only has a finite variety of transition states. For  $n \neq 0$ , these transitions are  $n \mapsto \{n - 1, n, n + 1\}$  and  $n \mapsto \{n, n + 1\}$ , for  $n = 0$ .

Considering this brief discussion, the dynamics described by Eq. (8) and the imposed conditions represents a time homogeneous *birth-death process*. A naive way to picture the process in the context of *queueing theory* [14] is through one queue representing all ancestors waiting to be classified if they are blood related or not by one server.

In Appendix D, we show how to choose a numerical matrix  $\mathbf{Q}$ . Finally, the evolution equations takes the form

$$\begin{aligned} d_t P_n(t) &= P_{n+1}(t) - 2P_n(t) + P_{n-1}(t), \\ d_t P_0(t) &= P_1(t) - P_0(t), \end{aligned} \quad (10)$$

and together with the initial condition, which is  $P_n(0) = \delta_{n0}$ , we obtain the explicit solution [14]

$$P_n(t) = e^{-2t} [I_n(2t) + I_{n+1}(2t)], \quad (11)$$

where  $I_n(x)$  is the modified Bessel function [15]. A brief description to obtain the solution Eq. (11) is also present in Ref. [14]. There is a construction of the generatrix function  $g(t, z) = \sum_{n \in \mathbb{N}} P_n(t) z^n$ ; see Eq. (D2), where from Eq. (8) we derive an equation for  $g(t, z)$ .

Equation (10) is the generic expression for all Markov processes on denumerable sample spaces and continuous time with a particular value of  $\mathbf{Q}$ .

### IV. THE GAUGED DISTRIBUTION OF ANCESTORS

As we have previously argued, before using this distribution to calculate the moments, it is necessary to perform a renormalization process. The reason is that the solution given

by Eq. (11) is normalized over  $\mathbb{N}$ . We perform a gauge transformation [16], denoted by  $\mathfrak{g}_t$ , which is applied to the probability distributions as

$$\mathfrak{g}_t : P_n(t) \longrightarrow \lambda(t) P_n(t). \quad (12)$$

The transformation Eq. (12) leaves the evolution Eq. (8) invariant and allows both distributions to describe a Markov process. We denote  $\mathfrak{p}_n(t) = \lambda(t) P_n(t)$  as the gauge-transformed distribution of  $P_n(t)$ . The action of the group  $\mathfrak{g}_t$  applied to the distribution  $P_n(t)$  leads to a distribution  $\mathfrak{p}_n(t)$  defined over  $\mathbb{S}_t$ . This idea can be understood in the context of conditional probabilities, with which we can obtain a projection of the distribution on  $\mathbb{N}$  into  $\mathbb{S}_t$ , keeping the correct normalization.

In other words, the transformation  $\mathfrak{g}_t$  leads to a new random variable  $X$ , which is the *gauge transformation* of  $X$ .

To preserve the invariance of Eq. (8) under  $\mathfrak{g}_t$ , we introduce a covariant derivative

$$D_t = d_t - \omega(t), \quad (13)$$

where  $\omega(t) = d_t \lambda(t) [\lambda(t)]^{-1}$ . See the Appendix for a more extensive explanation.

The expectation value of  $X$  raised to a positive integer power  $k$  is  $\langle X^k(t) \rangle = \sum_n n^k \mathfrak{p}_n(t)$ . This allows us to write a general relation between  $\langle X^k(t) \rangle$  and  $\langle X^k(t) \rangle$ :

$$\langle X^k(t) \rangle = \lambda(t) \langle X^k(t) \rangle. \quad (14)$$

Rescaling the process described by  $X(t)$  and using the solution Eq. (11), we calculated the first two cumulants:

$$\begin{aligned} \langle X(t) \rangle &= \lambda(t) \langle X(t) \rangle, \\ \langle [X(t) - \langle X(t) \rangle]^2 \rangle &= \lambda(t) [2t - \langle X(t) \rangle - \langle X(t) \rangle^2], \end{aligned} \quad (15)$$

where

$$\langle X(t) \rangle = e^{-2t} \left[ 2t I_1(2t) + \left( 2t + \frac{1}{2} \right) I_0(2t) \right] - \frac{1}{2}. \quad (16)$$

From Eq. (2) the variance of  $x$  is equal to the variance of  $y$ . The same argument is valid for  $X$  and  $Y$ .

We define the *standard deviation* of  $Y(t) = 2^{t+1} - X(t)$ , denoted by  $\sigma(t)$ , as the square root of the second equation of Eqs. (15), which quantifies the statistical fluctuation.

As we consider a constant function  $\omega(t)$ , then

$$\lambda(t) = 2^{at+b}. \quad (17)$$

We have obtained a family of functions for the expectation number of ancestors,

$$\alpha(t) = 2^{t+1} - \lambda(t) \langle X(t) \rangle, \quad (18)$$

parametrized by the real numbers  $a$  and  $b$  of Eq. (17).

If the expected value satisfies  $\alpha(t_1) = \alpha_1$  and  $\alpha(t_2) = \alpha_2$ , for two generations  $t_1$  and  $t_2$  such that  $t_1 \neq 0 \neq t_2$ , the parameters  $a$  and  $b$  can be obtained by

$$a = \frac{1}{t_2 - t_1} \log_2 \left[ \frac{2^{t_2+1} - \alpha_2 \langle X(t_1) \rangle}{2^{t_1+1} - \alpha_1 \langle X(t_2) \rangle} \right], \quad (19)$$

$$b = \frac{1}{t_2 - t_1} \left\{ t_2 \log_2 \left[ \frac{2^{t_1+1} - \alpha_1}{\langle X(t_1) \rangle} \right] - t_1 \log_2 \left[ \frac{2^{t_2+1} - \alpha_2}{\langle X(t_2) \rangle} \right] \right\},$$

where naturally  $\alpha_i \leq 2^{t_i+1}$ , for  $i = 1, 2$ , to ensure good definition of  $a$  and  $b$ .

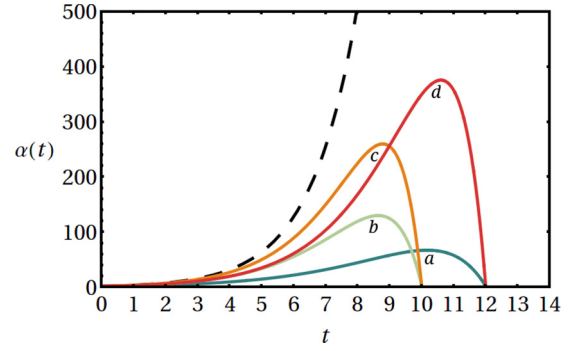


FIG. 2. (Color online) The four values of  $\{a, b\}$  are obtained from (19) and parametrized from  $(t_1, \alpha_1, t_2, \alpha_2) = (t_1, \xi 2^{t_1+1}, T, 2)$ , where  $0 \leq \xi \leq 1$ . In  $t_1$  the curve reaches the fraction  $\xi$  of the total number of possible ancestors for that generation, while  $t_2$  defines the maximum generation range denoted by  $T$ . The geometric progression  $2^{t+1}$  is in dash black line. The  $\{a, b, c, d\}$  (blue, green, orange, red) lines can be obtained respectively from  $(t_1, \alpha_1, t_2, \alpha_2) \in \{(3, 0.4 \times 2^4, 12, 2), (3, 0.7 \times 2^4, 10, 2), (3, 0.9 \times 2^4, 10, 2), (3, 0.7 \times 2^4, 12, 2)\}$ .

The gauge transformation modulates the amplitude of  $\langle X(t) \rangle$ . This allows us to define the notion of *horizontal* and *vertical range* of  $\alpha(t)$ . One important point of the curve  $\alpha(t)$  is the *maximum generation range*, this is a nonzero generation  $T$  in which  $\alpha$  becomes equal to 2. Another interesting point is the maximum of  $\alpha(t)$ , which determines the *intensity of the process*. Without loss of generality we can choose  $t_2 = T$ , in which case  $\alpha(t_2) = 2$ , and  $\alpha(t_1) = \sup\{\alpha(t) : t \in [0, T]\}$ . For any pair of different points,  $(t_1, \alpha_1)$  and  $(t_2, \alpha_2)$ , considered relevant, we select one and only one curve of the family, parametrized by  $a$  and  $b$  given by Eq. (19). The gauge transformation  $\mathfrak{g}_t$ , through the  $a$  and  $b$  parameters, controls both *horizontal range* and *vertical range* of the process. This  $T$  may not be a realistic value, but fix a maximum number of generations a particular species may have.

For illustrative purposes, in Fig. 2 we have selected four curves to the expectation number of ancestors  $\alpha(t)$  given by Eq. (18) and parametrized by different values of  $\{a, b\}$ . We include a geometric progression  $2^{t+1}$ , which corresponds to the maximum possible number of ancestors in each  $t$  generation.

Figure 3 shows three realizations of the number of ancestors in terms of the expectation value  $\alpha(t)$  and a measure of the dispersion given by  $\sigma(t)$ , for particular values of  $a$  and  $b$ .

This model may be employed in order to recognize a possible threshold to identify high endogamic populations as well as its possible causes. Using the genealogical tree, the model can be used to indicate which living species may be near to extinction.

## V. FINAL COMMENTS AND POSSIBLE MODEL EXTENSIONS

The model explained above allows us to calculate the expectation number of ancestors in each generation, considering the possibility of blood relationship between individuals of the same generation and a population of ancestors of maximum size  $2^{t+1}$ . But there are two possible generalizations. The model can be extended to take into account relationships between individuals of other adjacent generations using a



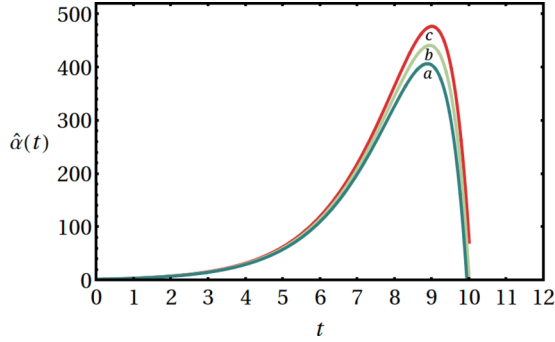


FIG. 3. (Color online) A band of curves, defined by the set  $\mathcal{B} = \{\hat{\alpha}(t) : \hat{\alpha}(t) \in [\alpha(t) - \sigma(t), \alpha(t) + \sigma(t)]\}$ , contains the expectation value  $\alpha(t)$  given by the equation (18) and its statistical fluctuation at  $1-\sigma$ . The  $\{a, b, c\}$  (blue, green, red) lines corresponds to the realization of  $\{\alpha(t) - \sigma(t), \alpha(t), \alpha(t) + \sigma(t)\}$ , with the values of  $(t_1, \alpha_1, t_2, \alpha_2) = (6, 0.9 \times 2^7, 10, 2)$ , respectively.

similar idea, simply considering higher-order Markov chains. By introducing the corresponding terms in the infinitesimal generator  $\mathbf{Q}$ , for example an absorbing barrier [17], the same model can be used to calculate the expectation number of ancestors in a specific population with additional restrictions such as isolation, immigration, specific reproductive behavior, or cultural restrictions for the human case.

In future applications we could generalize the model through a new constraint to fix the maximum number of individuals at a certain generation. This proposal implies a generalization of this work in which the maximum number of ancestors will be given by a piecewise function  $\gamma(t)$  instead of  $2^{t+1}$ . This leads us to slightly modify the process defined at the beginning in Eq. (2) as  $y(t) = \gamma(t) - x(t)$  and the endpoint of the sample space for  $x(t)$  becomes  $n_t = \gamma(t) - 2$ . This generalization includes a time inhomogeneity in the infinitesimal generator  $\mathbf{Q}$  and preserves an appropriate renormalization.

We can include these possible extensions using the process  $\{X(t)\}$  and study a most general gauge  $\mathfrak{g}_t$  transform given by a linear transformation,

$$\mathfrak{g}_t : \{P_m(t)\}_{m \in \mathbb{N}} \longrightarrow \mathfrak{p}_n(t) = \sum_{m \in \mathbb{N}} \lambda_{nm}(t) P_m(t), \quad (20)$$

where  $\lambda = \{\lambda_{nm}\}$  is a nonsingular matrix. We can express the last expression in a matrix form,

$$\mathfrak{g}_t : \boldsymbol{\phi} \longrightarrow \boldsymbol{\phi}' = \lambda \boldsymbol{\phi}, \quad (21)$$

where  $\boldsymbol{\phi}'(t) = [p_0(t), p_1(t), \dots]^T$ .

In order to preserve the invariance of Eq. (8) under this generalization of  $\mathfrak{g}_t$ , we introduce the corresponding covariant derivative:

$$D_t = d_t - \boldsymbol{\omega}(t), \quad (22)$$

where  $\boldsymbol{\omega}(t) = d_t \lambda(t) [\lambda(t)]^{-1}$ .

The evolution equation for  $\boldsymbol{\phi}'(t)$  is also invariant under the local dilation group,

$$D_t \boldsymbol{\phi}'(t) = \mathbf{Q}'(t) \boldsymbol{\phi}'(t), \quad (23)$$

and the gauged infinitesimal generator is now

$$\mathbf{Q}'(t) = \lambda(t) \mathbf{Q} [\lambda(t)]^{-1}, \quad (24)$$

which corresponds to a similarity transformation of  $\mathbf{Q}$ .

This model can be applied to describe other biological or physical systems with similar dynamics. Statistical models of biparental reproduction have already been compared with physical systems, such as spin-glass systems [18]. In this regard the *evolutionary graph theory* is an approach to study how topology affects the evolution of a population [19].

Other analogous processes to the biparental reproduction in physics are described with similar statistical or Markovian models [10]. In high-energy physics the production of a cascade by a cosmic ray is described by the Heitler model [20]. Although this model is different from the one presented here, we could compare the number of ancestors with the number of particles in each generation and reinterpret the results in terms of these kinds of phenomena.

It is possible to estimate the maximum generation range,  $T$ , searching in the fossil record the first time that a particular species appears and use its reproductive rate. In this way we are classifying each species not in terms of lifetime on Earth (time units), but according to the notions of generational patterns.

The interaction of various processes can be combined with universal common ancestor's models [21] to understand the development of a certain species. The ideas in the current model can be used in biology, population ecology, and genetics. An important achievement of the model is that based on the previous knowledge of the lifetime of a certain species, we can calculate the number of ancestors in each generation of this species.

More fundamental uses of this idea can be found in mathematics, in particular in theory of stochastic processes and in the physics area connected with the theory of stochastic processes. Future research through a Lagrangian description may find novel applications of the present proposal. In this case we will consider the probabilities  $\{p_n(t)\}$  as the set of generalized coordinates.

## ACKNOWLEDGMENTS

We thank our respective Ph.D. advisors: Fernando Cornet and María Teresa Dova, Hernán Wahlberg. We thank Carlos García Canal and Huner Fanchiotti for their helpful criticism. We are indebted to Federico Agnolin and Ben Page for reading the manuscript and providing advice. We also thank our anonymous reviewers for their contribution to this work. And finally, a special mention to Micaela Moretton, María Clara Caruso, and Gabriel Lio for local support.

## APPENDIX A: ON THE EVOLUTION EQUATION IN S

We defined the random variable  $y(t)$  associated with the number of ancestors as

$$y(t) = 2^{t+1} - x(t), \quad (A1)$$

where  $\mathbb{S}_t$  is the sample space of  $x(t)$  and  $t$  is a continuous variable. Then, according to a discretization process, the distribution of ancestors is obtained and the variable  $t$  will be the number of generations.

We denote by  $\alpha(t)$  the expectation number of  $y(t)$  as

$$\alpha(t) = 2^{t+1} - \langle x(t) \rangle, \quad (\text{A2})$$

where  $\langle x(t) \rangle$  constrains the number of ancestors under the blood-relationship hypothesis. We will focus on the process of this random variable  $x(t)$  and our probability distribution denoted by  $p_n(t) = \mathbb{P}[x(t) = n]$ .

In the problem of counting ancestors the sample space  $\mathbb{S}_t$  is different for each  $t$  generation. For this reason we introduced a fictitious process on a larger sample space  $\mathbb{S}$  than the original  $\mathbb{S}_t$ . Therefore, we have solved the problem in  $\mathbb{S}$  and renormalized the distribution to take into account the interaction with the lost boundary of  $\mathbb{S}_t$ . In other words, we have considered a *dilution* of  $\mathbb{S}_t$  into  $\mathbb{S}$ . In particular, we used the set of natural numbers  $\mathbb{S} = \mathbb{N}$ , including the 0 element.

We considered the study of the evolution over a generic denumerable sample space  $\mathbb{S} = [0, N] \subset \mathbb{N}$  with the random variable  $X(t)$  and probability distribution  $P_n(t) = \mathbb{P}[X(t) = n]$ . This evolution is governed by the conditional probability given by

$$\mathcal{P}_{nm}(t, s) = \mathbb{P}[X(t) = n | X(s) = m]. \quad (\text{A3})$$

The matrix  $\mathcal{P}(t, s)$  satisfies the *Chapman-Kolmogorov* equation [11, 17],

$$\mathcal{P}(t, s) = \mathcal{P}(t, u)\mathcal{P}(u, s), \quad (\text{A4})$$

for  $0 \leq s \leq u \leq t$ . Also the sum of the elements of each column is

$$\sum_{n \in \mathbb{S}} \mathcal{P}_{nm}(t, s) = 1. \quad (\text{A5})$$

For the general case we develop a power series of the matrix  $\mathcal{P}(t + \epsilon, s)$ ; for a fixed value of  $s$ , we have

$$\mathcal{P}_{nm}(t + \epsilon, s) = \mathcal{P}_{nm}(t, s) + \epsilon \partial_t \mathcal{P}_{nm}(t, s) + \dots, \quad (\text{A6})$$

where  $\partial_t$  is a simplified notation of partial time derivative  $\frac{\partial}{\partial t}$ .

In order to obtain Eq. (4) we study the time evolution  $t \mapsto t + \epsilon$ , for small value of  $\epsilon$ . We need to know  $\mathcal{P}_{nm}(t + \epsilon, t)$  then Eq. (A6) becomes

$$\mathcal{P}_{nm}(t + \epsilon, t) = \delta_{nm} + \epsilon \partial_t \mathcal{P}_{nm}(t, s)|_{s=t} + \dots \quad (\text{A7})$$

We recognize the second term of Eq. (A.8) as the infinitesimal generator  $\mathcal{Q}_{nm}(t)$ :

$$\mathcal{Q}_{nm}(t) = \lim_{\epsilon \rightarrow 0} \frac{\mathcal{P}_{nm}(t + \epsilon, t) - \delta_{nm}}{\epsilon}. \quad (\text{A8})$$

We assumed that  $t$  is continuous. This allows us to evaluate the process at any  $t$  between two generations, but it also reduces the number of possible states in an infinitesimal evolution. Therefore, only transitions to the nearest states are allowed. For the ancestry problem, the infinitesimal time evolution has a finite number of transition states. These transitions are denoted by  $n \mapsto n'$ , where  $n' \in T_n$ ; i.e.,  $n'$  depends on the initial state  $n$ . We write explicitly  $T_0 = \{0, 1\}$ ,  $T_N = \{N - 1, N\}$  and for  $n \neq 0, N$ :  $T_n = \{n - 1, n, n + 1\}$ . Furthermore, if  $|n' - n| > 1$  the corresponding transition probability is zero.

From Eq. (A5) the matrix  $\mathcal{P}(t + \epsilon, t)$  is normalized for all  $t$  and small  $\epsilon$  as

$$\sum_{n' \in T_n} \mathcal{P}_{n'n}(t + \epsilon, t) = 1. \quad (\text{A9})$$

Note that  $n'$  runs over  $T_n$ , depending on whether  $n$  is equal to 0,  $N$ , or any other value of  $\mathbb{S} - \{0, N\}$ .

We express  $\mathcal{P}_{n'n}(t + \epsilon, t)$  for these three cases from Eq. (A.8):

$$\begin{aligned} \mathcal{P}_{n-1n}(t + \epsilon, t) &= \mu_n(t)\epsilon + \mathcal{O}_t(\epsilon), \\ \mathcal{P}_{n+1n}(t + \epsilon, t) &= \nu_n(t)\epsilon + \mathcal{O}_t(\epsilon), \\ \mathcal{P}_{n'n}(t + \epsilon, t) &= 0, \quad |n' - n| > 1, \\ \mathcal{P}_{nn}(t + \epsilon, t) &= 1 - [\nu_n(t) + \mu_n(t)]\epsilon + \mathcal{O}_t(\epsilon), \end{aligned} \quad (\text{A10})$$

where  $\mathcal{O}_t(x)$  represents a type of function that goes to zero with  $x$  faster than  $x$ , for a given  $t$ , that is

$$\lim_{x \rightarrow 0} \frac{\mathcal{O}_t(x)}{x} = 0. \quad (\text{A11})$$

The fourth equation of Eqs. (A10) is obtained through the first three of them. In the general case we proceed as follows from Eq. (A5),

$$\mathcal{P}_{nn}(t, s) = 1 - \sum_{n' \in \mathbb{S} - \{n\}} \mathcal{P}_{n'n}(t, s), \quad (\text{A12})$$

then from an infinitesimal time evolution  $s \equiv t \mapsto t + \epsilon$  and Eq. (A9),

$$\begin{aligned} \mathcal{P}_{nn}(t + \epsilon, t) &= 1 - \sum_{n' \in T_n - \{n\}} \mathcal{P}_{n'n}(t + \epsilon, t), \\ \mathcal{P}_{nn}(t + \epsilon, t) &= 1 - \mathcal{P}_{n+1n}(t + \epsilon, t) - \mathcal{P}_{n-1n}(t + \epsilon, t), \end{aligned}$$

which is the fourth equation of Eqs. (A10).

Replacing Eqs. (A10) in Eq. (4) and written for  $n \neq 0$ , we have

$$\begin{aligned} P_n(t + \epsilon) &= [\nu_{n-1}(t)\epsilon + \mathcal{O}_t(\epsilon)] P_{n-1}(t) \\ &\quad + [\mu_{n+1}(t)\epsilon + \mathcal{O}_t(\epsilon)] P_{n+1}(t) \\ &\quad + [1 - \nu_n(t)\epsilon - \mu_n(t)\epsilon + \mathcal{O}_t(\epsilon)] P_n(t), \end{aligned}$$

then

$$\begin{aligned} \frac{P_n(t + \epsilon) - P_n(t)}{\epsilon} &= \left[ \nu_{n-1}(t) + \frac{\mathcal{O}_t(\epsilon)}{\epsilon} \right] P_{n-1}(t) \\ &\quad + \left[ \mu_{n+1}(t) + \frac{\mathcal{O}_t(\epsilon)}{\epsilon} \right] P_{n+1}(t) \\ &\quad - \left[ \nu_n(t) + \mu_n(t) + \frac{\mathcal{O}_t(\epsilon)}{\epsilon} \right] P_n(t), \end{aligned}$$

taking the limit  $\epsilon \rightarrow 0$ ,

$$\begin{aligned} d_t P_n(t) &= \nu_{n-1}(t) P_{n-1}(t) \\ &\quad + \mu_{n+1}(t) P_{n+1}(t) \\ &\quad - [\nu_n(t) + \mu_n(t)] P_n(t). \end{aligned} \quad (\text{A13})$$

The stochastic process described by Eq. (A13) corresponds to the general class of stochastic dynamics called *birth and death process*, which includes the *queueing process* [11, 14].

The functions  $\mu_n(t)$  and  $\nu_n(t)$  are part of the infinitesimal generator  $\mathbf{Q}(t)$ . From the first equation of Eqs. (A10) we have

$$\mu_n(t) = \lim_{\epsilon \rightarrow 0} \frac{\mathcal{P}_{n-1n}(t + \epsilon, t)}{\epsilon}, \quad (\text{A14})$$

which is exactly the element  $Q_{n-1n}(t)$  of Eq. (A.8).

In summary, we list all the elements of  $\mathbf{Q}(t)$ :

$$\begin{aligned} Q_{nn}(t) &= -[\nu_n(t) + \mu_n(t)], & Q_{n-1n}(t) &= \mu_n(t) \\ Q_{n+1n}(t) &= \nu_n(t), & Q_{n'n}(t) &= 0, \quad |n' - n| > 1. \end{aligned} \quad (\text{A15})$$

We write Eq. (A13) in a matrix form as

$$d_t \boldsymbol{\phi}(t) = \mathbf{Q}(t) \boldsymbol{\phi}(t), \quad (\text{A16})$$

where  $\boldsymbol{\phi}(t) = [P_0(t), \dots, P_N(t)]^\top$  and  $\mathbf{Q}(t) = \{Q_{n'n}(t)\}$  is given by Eq. (A15).

It should also be pointed out that the coefficients  $\mu_0$  and  $\nu_N$  must be zero; otherwise, we require more states than  $[0, N]$  in  $\mathbb{S}$ . In other words, if  $\mu_0$  or  $\nu_N$  are not equal to zero, the left side of Eq. (A9) is not equal to one.

## APPENDIX B: HOMOGENEOUS HYPOTHESIS

In this section we show how the hypothesis of spatial and temporal homogeneity are used working in the dilated sample space.

We have already said that if there exists a certain  $T$  generation that can be considered as the stop of the process, the upper limit of the dilated space  $\mathbb{S}$ , denoted by  $N$ , can be chosen as

$$N = n_T, \quad (\text{B1})$$

where  $n_T = \sup\{n : n \in \mathbb{S}_t, \forall t \in [0, T]\}$ .

Also, we considered a dilution of  $\mathbb{S}_t$  into  $\mathbb{N}$ , i.e., the generic dilated space  $\mathbb{S}$  is equal to  $\mathbb{N}$ , or  $N \rightarrow \infty$ . This assumption is true from  $\mathbb{S}_t \subseteq \mathbb{N}$ , no matter how big  $n_T$  is. In this case, the space-time on the process is infinite and we have an infinitesimal generator on  $\mathbb{N}$ , independent of the state of the random variable  $X$  and time-independent, and the process is *space-time homogeneous*.

However, the effect of these hypotheses can be compensated with an interaction with the process on  $\mathbb{N} - \mathbb{S}_t$ ; see Appendix D for more details.

Essentially, we will say that the renormalized distribution defined on  $\mathbb{N}$  is equivalent to the distribution defined on  $\mathbb{S}_t$ . This equivalence is based on the invariance of evolution equation. In this way both distributions correspond to a Markov process; see Appendix D.

## APPENDIX C: ON THE SPACE-TIME HOMOGENEOUS SOLUTION IN $\mathbb{N}$

The evolution equation in the dilated sample space  $\mathbb{N}$  under the space-time homogeneity is

$$d_t \boldsymbol{\phi}(t) = \mathbf{Q} \boldsymbol{\phi}(t), \quad (\text{C1})$$

with  $\boldsymbol{\phi}(t) = [P_0(t), P_1(t), \dots]$  and identify the matrix  $\mathbf{Q}$  as

$$\mathbf{Q} = \begin{pmatrix} -\nu & \mu & 0 & \dots \\ \nu & -\mu - \nu & \mu & \dots \\ 0 & \nu & -\mu - \nu & \dots \\ 0 & 0 & \nu & \dots \\ 0 & 0 & 0 & \dots \\ \vdots & \vdots & \vdots & \ddots \end{pmatrix}. \quad (\text{C2})$$

Under the initial condition  $\boldsymbol{\phi}(0) = (1, 0, \dots)^\top$ , the solution of Eq. (C1) with Eq. (C2) is

$$\begin{aligned} P_n(t) &= e^{-(\nu+\mu)t} \left[ \rho^{n/2} I_n(\zeta t) + \rho^{(n-1)/2} I_{n+1}(\zeta t) \right. \\ &\quad \left. + (1 - \rho) \rho^n \sum_{j=n+2}^{\infty} \rho^{-j/2} I_j(\zeta t) \right] \end{aligned} \quad (\text{C3})$$

where  $I_n(x)$  is the modified Bessel function [15],  $\rho = \nu/\mu$  and  $\zeta = 2\sqrt{\nu\mu}$ . The first solution of Eq. (C1) appeared in the 1950s; see Refs. [22–25]. A description to obtain the solution Eq. (C3) is also presented in Ref. [14].

## APPENDIX D: MOMENTS OF THE DISTRIBUTION

In this section we calculate the first two cumulants of the distribution obtained in Eq. (C3).

First of all we demonstrated the existence of all  $k$  moments of the distribution  $P_n(t)$  defined by

$$\langle X^k(t) \rangle = \sum_{n \in \mathbb{N}} n^k P_n(t), \quad (\text{D1})$$

with  $k \in \mathbb{N}$ . It is possible to demonstrate that all series defined above converge uniformly  $\forall t$ . To demonstrate this, we define the *generatrix function*

$$g(t, z) = \sum_{n \in \mathbb{N}} P_n(t) z^n, \quad (\text{D2})$$

for  $z \in \mathbb{R}$ . If it converges,  $g(t, z)$  is well defined. For our case, we know that

$$\sum_{n \in \mathbb{N}} I_n(x) \rightarrow \frac{1}{2} [e^x + I_0(x)] \quad (\text{D3})$$

converges uniformly [15]. This allows us to write the distribution's norm and demonstrate that it converges uniformly:

$$\sum_{n \in \mathbb{N}} P_n(t) \rightarrow 1. \quad (\text{D4})$$

Another argument for the general birth-death process, based on the nature of the coefficients  $\{(v_n, \mu_{n+1}) : n \in \mathbb{N}\}$ , leads us to the same conclusion [17].

For each  $t$ , we demonstrated that  $g(t, 1) \rightarrow 1$  uniformly, then for the Abel's theorem [26],

$$\sum_{n \in \mathbb{N}} P_n(t) z^n \rightarrow g(t, z), \quad (\text{D5})$$

uniformly for each  $z \in [0, 1]$ .

Therefore, Eq. (D2) can be derived term by term keeping the uniform convergence  $\forall(t, k)$ :

$$\sum_{n \in \mathbb{N}} \frac{n!}{(n-k)!} P_n(t) \longrightarrow \partial_z^k g(t, z)|_{z=1}. \quad (\text{D6})$$

Finally,  $\langle X^k(t) \rangle$  can be obtained as a combination of the series given by Eq. (D6), which converges to  $\partial_z^k g(t, z)|_{z=1}$ , and this completes the demonstration.

We can use two methods for the calculation of the first two moments, given the solution Eq. (C3).

The first one is by definition Eq. (D1) and using the identity [15]

$$nI_n(x) = \frac{x}{2}[I_{n-1}(x) - I_{n+1}(x)]. \quad (\text{D7})$$

The second one is based on the series Eq. (D1) converging uniformly. We can derive term by term the series Eq. (D1). Using the evolution Eq. (C1) with Eq. (C2), we can obtain a differential equation for the expectation value  $\langle X(t) \rangle$  and  $\langle X^2(t) \rangle$ :

$$d_t \langle X(t) \rangle = (v - \mu) + \mu P_0(t), \quad (\text{D8})$$

$$d_t \langle X^2(t) \rangle = 2v - d_t \langle X(t) \rangle + 2(v - \mu) \langle X(t) \rangle. \quad (\text{D9})$$

For the initial condition  $P_n(0) = \delta_{n0}$  then  $\langle X^k(0) \rangle = 0$  and integrate the last two equations:

$$\langle X(t) \rangle = (v - \mu)t + \mu \int_0^t P_0(\tau) d\tau, \quad (\text{D10})$$

$$\langle X^2(t) \rangle = 2vt - \langle X(t) \rangle + 2(v - \mu) \int_0^t \langle X(\tau) \rangle d\tau. \quad (\text{D11})$$

Equation (D10) shows that to determine  $\langle X(t) \rangle$  it is sufficient to know the distribution of probability of no blood relationship  $P_0(t)$  from Eq. (C3). Also, Eq. (D11) shows that to determine  $\langle X^2(t) \rangle$  it is sufficient to know  $\langle X(t) \rangle$ .

These two methods arrive at the same result for  $\langle X(t) \rangle$ .

We are interested in computing the first two cumulants. The first one is the expectation value and the second one is defined as a function of the first two moments by  $\langle [X(t) - \langle X(t) \rangle]^2 \rangle = \langle X^2(t) \rangle - \langle X(t) \rangle^2$ , respectively.

We have said that this is due to dilation of the sample space  $S_t \mapsto \mathbb{N}$ . This implies that distributions defined on  $\mathbb{N}$  takes smaller values than they should take on  $S_t$ .

Also, we show how  $\langle X(t) \rangle$  is small compared to  $2^{t+1}$ , for any value of  $v \geq 0$  and  $\mu \geq 0$ .

Using Eq. (D10) and  $0 \leq P_0(t) \leq 1$ , we can express

$$|\langle X(t) \rangle| \leq |v - \mu|t + \mu \int_0^t P_0(\tau) d\tau \leq |v - \mu|t + \mu t. \quad (\text{D12})$$

This shows that  $\langle X(t) \rangle$  is subordinated to a linear function in  $t$ , for all  $\mu, v$ , then the expected value of ancestors  $\langle Y(t) \rangle = 2^{t+1} - \langle X(t) \rangle$  will grow indefinitely with  $t$  as  $2^{t+1}$ . We can consider by ignorance  $v = \mu$ , in the sense of not knowing the functional form of the trend of the number of ancestors. Although the knowledge of any particular trend can be introduced in the gauge transformation. And, finally,

without loss of generality we can take  $\mu = 1 = v$ , since the problem of nonsaturation will be solved by dilating the distribution to compensate for the dilution of the sample space, as we shall see in the next section. This dilation can be seen as a renormalization. In the first place, we considered the case where  $X(t)$  is defined on  $\mathbb{N}$  with no renormalization at all. In this way, the renormalization is interpreted as an operation where the correct scale of the interaction is retrieved modifying  $\langle X(t) \rangle$ , as if we had solved the problem in the original sample space  $S_t$ .

## APPENDIX E: ON THE INTERACTION WITH A FICTITIOUS ENVIRONMENT

We analyzed one of the main concepts: the dilution of the sample space involves the study of an interaction between the initial sample space with the fictitious environment. The introduction of the dilution takes into account the interaction with the lost boundary of  $S_t$ , under the condition that  $S$  is large enough to include  $S_t$  for all  $t$ .

We can view the process on  $S_t$  as the result of a process on  $S$  that interacts with another process defined on the complement set  $\mathbb{R}_t = S - S_t$ . This point of view can be described in a mathematical precise sense. We define the associated vector spaces  $\{S, S_t, R_t\}$  to the sets  $\{S, S_t, \mathbb{R}_t\}$ , where  $\phi$  is a vector in  $S$ , which  $\dim(S) = N + 1$ . If the sample space  $S_t$  has  $n_t + 1$  number of states, we define  $\varphi$  as the first  $n_t + 1$  component of  $\phi$ , i.e.,  $\varphi$  is vector of  $S_t$ . We expressed these vectors in a canonical basis  $\{e_n\}_{n=0, \dots, N}$ , such that

$$\begin{aligned} \phi(t) &= \sum_{n \in S} P_n(t) e_n \\ \phi(t) &= \sum_{n \in S_t} P_n(t) e_n + \sum_{n \in \mathbb{R}_t} P_n(t) e_n, \end{aligned} \quad (\text{E1})$$

then we have

$$\phi(t) = \varphi(t) + \psi(t), \quad (\text{E2})$$

where  $e_0 = (1, 0, \dots, 0)^T$ ,  $e_1 = (0, 1, \dots, 0)^T$ , and so forth. For construction,  $S_t$  is orthogonal to  $R_t$ . The dimensions of these vector spaces are determined by  $\dim(S) = N + 1$  and  $\dim(S_t) = n_t + 1$ .

From Eq. (E2), we see, roughly speaking, that the process on  $S_t$  is the result of interaction between the process on  $S$  and  $\mathbb{R}_t$  through

$$\varphi(t) = \phi(t) - \psi(t). \quad (\text{E3})$$

We can write down the evolution Eq. (A16) in the form

$$d_t \varphi(t) = \mathbf{Q}_{ss}(t) \varphi(t) + \mathbf{Q}_{sr}(t) \psi(t), \quad (\text{E4})$$

$$d_t \psi(t) = \mathbf{Q}_{rs}(t) \varphi(t) + \mathbf{Q}_{rr}(t) \psi(t), \quad (\text{E5})$$

where  $\mathbf{Q}_{ab}$  is the  $a \times b$  block matrix of  $\mathbf{Q}$ , for  $a, b \in \{s, r\}$ ,  $s = n_t + 1$ , and  $r = N - n_t$ , explicitly

$$\mathbf{Q} = \begin{pmatrix} \mathbf{Q}_{ss} & \mathbf{Q}_{sr} \\ \mathbf{Q}_{rs} & \mathbf{Q}_{rr} \end{pmatrix}. \quad (\text{E6})$$

We see that the equations for  $\varphi$  and  $\psi$  are coupled, hence the interaction character that was noted above. To be more



specific, we write a relation of two *partial solutions*  $\psi$  and  $\varphi$ ,

$$\psi(t) = \int_{\mathbb{R}} \mathbf{K}_{rs}(t, t') \varphi(t') dt', \quad (\text{E7})$$

where Eq. (E7) satisfies Eq. (E5) and  $\mathbf{K}_{rs}(t, t')$  is the kernel of this transformation defined as

$$\mathbf{K}_{rs}(t, t') = \mathcal{T} \left\{ \exp \left[ \int_{t'}^t \mathbf{Q}_{rr}(\tau) d\tau \right] \right\} \mathbf{Q}_{rs}(t') \theta(t - t'), \quad (\text{E8})$$

and  $\theta(t)$  are the *unit step* distribution and  $\mathcal{T}$  is the time-ordered operator defined as

$$\mathcal{T}[\mathbf{A}(t)\mathbf{A}(u)] = \begin{cases} \mathbf{A}(t)\mathbf{A}(u) & : t > u \\ \mathbf{A}(u)\mathbf{A}(t) & : u > t \end{cases} \quad (\text{E9})$$

We can obtain another expression for Eq. (E4),

$$d_t \varphi(t) = \int_{\mathbb{R}} \mathbf{K}_{ss}(t, t') \varphi(t') dt', \quad (\text{E10})$$

where  $\mathbf{K}_{ss}$  is the kernel of the integrodifferential Eq. (E10) defined as

$$\mathbf{K}_{ss}(t, t') = \mathbf{Q}_{ss}(t') \delta(t - t') + \mathbf{Q}_{sr}(t) \mathbf{K}_{rs}(t, t'), \quad (\text{E11})$$

and  $\delta(t)$  is the  $\delta$  distribution.

The same argument allows us to obtain an inverse relation of Eq. (E7),

$$\varphi(t) = \int_{\mathbb{R}} \mathbf{K}_{sr}(t, t') \psi(t') dt', \quad (\text{E12})$$

simply interchange in Eqs. (E8), (E10), and (E11) the quantities  $\varphi \longleftrightarrow \psi$ ,  $r \longleftrightarrow s$ .

The mathematical construction presented here shows how a process in  $\mathbb{S}$  can be described by the interaction of two subprocesses in  $\mathbb{S}_r$  and  $\mathbb{R}_r$ . Specifically, this interaction can be viewed in the relation Eq. (E7) or its inverse Eq. (E12).

In other words, the process on  $\mathbb{S}_r$  can be seen as the interaction between the processes on  $\mathbb{S}$  and  $\mathbb{R}_r = \mathbb{S} - \mathbb{S}_r$ ; this interaction emerges from the elements of infinitesimal transition probabilities present in  $\mathbf{K}_{ss}(t, t')$  through Eqs. (E11) and (E8).

## APPENDIX F: DILATION-DILUTION TRANSFORMATION

The *dilution* operation of  $\mathbb{S}_r$  into a larger sample space  $\mathbb{S}$  can also be understood as an *dilation* represented by the substitution rule  $n_r \mapsto N$ . In a general sense, every *dilution-dilation* transformation involves a renormalization of the distribution obtained above.

To illustrate this point let us consider two distributions,  $\{p_n : n \in \mathbb{S}\}$  and  $\{P_n : n \in \mathbb{S}\}$ , defined over the sample spaces  $\mathbb{s}$  and  $\mathbb{S}$ , respectively, such that  $\mathbb{s} \subset \mathbb{S}$  and both are normalized in each sample space. By definition we have

$$\sum_{n \in \mathbb{S}} P_n = \sum_{n \in \mathbb{S} - \mathbb{s}} P_n + \sum_{n \in \mathbb{s}} P_n = 1. \quad (\text{F1})$$

Since all terms of Eq. (F1) are positive, there exists a subset of  $\mathbb{s}$  in which the distribution  $P_n$ , restricted to  $\mathbb{s}$ , is smaller than

$\{p_n : n \in \mathbb{s}\}$ . We can see it in this from

$$\sum_{n \in \mathbb{S}} P_n < \sum_{n \in \mathbb{s}} p_n. \quad (\text{F2})$$

We want to describe the process  $\{p_n : n \in \mathbb{s}\}$  through the process of  $\{P_n : n \in \mathbb{S}\}$  *projecting* the distribution  $P_n$  over  $\mathbb{s}$ . Since there are values of  $\mathbb{s}$  for which  $P_n$  is less than  $p_n$ , that projection should amplify  $P_n$  to improve the stated description. This amplification corresponds to a dilation transformation.

We had mentioned that the renormalization is performed through a linear time-dependent transformation. As we conserve the linearity of Eq. (A16) and since the original sample space  $\mathbb{S}_r$  is time-dependent, the norm must be corrected locally. This *local* transformation is structured as a *gauge group*, specifically a *group of local dilations*.

On the other hand, this transformation can be viewed through a *projection* of the distribution  $P_n(t)$  on  $\mathbb{S}_r$ , which filter a subset of states and renormalizes the distribution. We used the concept of conditional probability to relate and motivate the renormalization through a gauge transformation. To be more clear, we distinguish the gauge transformed distribution  $p_n(t)$  and the projected distribution  $q_n(t)$ .

We used a dilation on a generic sample space  $\mathbb{S} = [0, N]$ , for a natural number  $N$ .

Let's consider the following example. If  $q_n(t)$  is defined as

$$q_n(t) = \mathbb{P}[X(t) = n | X(t) \in \mathbb{S}_r], \quad (\text{F3})$$

this distribution is equal to zero for all states such that  $n \in \mathbb{S} - \mathbb{S}_r$ . Applying the Bayes identity [17], we have

$$\mathbb{P}[A|B]\mathbb{P}[B] = \mathbb{P}[B|A]\mathbb{P}[A], \quad (\text{F4})$$

and we can express  $q_n(t)$  as

$$q_n(t) = \Lambda(t) P_n(t), \quad (\text{F5})$$

where

$$\Lambda(t) = \frac{\mathbb{P}[X(t) \in \mathbb{S}_r | X(t) = n]}{\mathbb{P}[X(t) \in \mathbb{S}_r]}. \quad (\text{F6})$$

The numerator of Eq. (F6) is only equal to 0 or 1 depending on whether  $n \in \mathbb{S} - \mathbb{S}_r$  or  $n \in \mathbb{S}_r$ , respectively. The denominator of Eq. (F6) is  $\mathbb{P}[X(t) \in \mathbb{S}_r] = \mathbb{P}[X(t) \leq n_r]$  and for the mutually exclusive events we have

$$\mathbb{P}[X(t) \in \mathbb{S}_r] = \sum_{n \in \mathbb{S}_r} P_n(t) < 1, \quad (\text{F7})$$

which implies  $q_n(t) \geq P_n(t)$ , for  $n \in \mathbb{S}_r$ . Equation (F5) involves a probability fraction that can be viewed as a projection operator that transforms distributions defined on  $\mathbb{S}$  into distributions defined on  $\mathbb{S}_r$ , by filtering selected states and redefining the correct normalization.

The last discussion about the conditional probability as a projection operation motivates the following construction. Given a process in  $\mathbb{S}$  described by  $\phi(t) = [P_0(t), \dots, P_N(t)]^T$ , which satisfies a general equation, similar to Eq. (A16),

$$d_t \phi(t) = \mathbf{Q}(t) \phi(t), \quad (\text{F8})$$

and the dilation transformation

$$g_t : P_n(t) \longrightarrow p_n(t), \quad (\text{F9})$$

where  $p_n(t) = \lambda(t)P_n(t)$  are the components of  $\phi'(t) = [p_0(t), \dots, p_N(t)]^T$ . We showed that the covariance requirement is satisfied if  $g_t$  is a *gauge transformation*, this implies the addition of a covariant derivative. To prove this statement we transform a general evolution Eq. (F8) through a local dilation group  $g_t$ :

$$d_t \phi'(t) - d_t \lambda(t) [\lambda(t)]^{-1} \phi'(t) = \mathbf{Q}(t) \phi'(t). \quad (\text{F10})$$

In order to keep the shape of Eq. (A16) when we transformed by  $g_t$ , we need to introduce an affine connection that can be implemented through a covariant derivative given by

$$D_t = d_t - \omega(t), \quad (\text{F11})$$

where  $\omega(t) = d_t \lambda(t) [\lambda(t)]^{-1}$  is the *gauge function*.

In this case, the evolution equation for  $\phi'(t)$  is

$$D_t \phi'(t) = \mathbf{Q}(t) \phi'(t), \quad (\text{F12})$$

which is invariant under the local dilation group.

The definition Eq. (F11) allows that the distributions  $P_n(t)$  and  $p_n(t)$ , related by gauge, corresponds to a Markov process.

For the case of this work, the gauge function  $\omega(t)$  is independent of the  $t$  generation, then  $\lambda(t) = 2^{at+b}$ , for  $a, b$  are real constants. Otherwise, the time dependence of  $\omega(t)$  introduces a time inhomogeneity in the process.

The example used to motivate the definition Eq. (F9) can be extended to describe a rich variety of cases. If we generalized  $q_n(t)$  as

$$q_n(t) = \mathbb{P}[X(t) = n | X(t) \in \mathbb{S}_t, Z], \quad (\text{F13})$$

where  $Z$  is an extra condition and eventually space-time dependent. The distribution Eq. (F13) is equal to zero for

$n \in \mathbb{S} - \mathbb{S}_t$ . Then, for the Bayes identity Eq. (F4) applied to Eq. (F13),

$$q_n(t) = \Lambda_n(t) P_n(t), \quad (\text{F14})$$

where

$$\Lambda_n(t) = \frac{\mathbb{P}[X(t) \in \mathbb{S}_t, Z | X(t) = n]}{\mathbb{P}[X(t) \in \mathbb{S}_t, Z]}. \quad (\text{F15})$$

In a more abstract sense, we consider a partition of  $\mathbb{S} = \{B_n : n \in \mathbb{S}\}$ , and let us consider the event  $A_n$  written in the following way:

$$A_n = \bigcup_{m \in \mathbb{S}} A_n \cap B_m. \quad (\text{F16})$$

Therefore, we use the law of total probability,

$$\mathbb{P}[A_n] = \sum_{m \in \mathbb{S}} \mathbb{P}[A_n | B_m] \mathbb{P}[B_m], \quad (\text{F17})$$

for a particular case of the partition  $B_m = \{m\}$  and  $\mathbb{P}[A_n] = q_n(t)$ , then

$$q_n(t) = \sum_{m \in \mathbb{S}} \Lambda_{nm}(t) P_m(t), \quad (\text{F18})$$

where

$$\Lambda_{nm}(t) = \mathbb{P}[A_n | X(t) = m]. \quad (\text{F19})$$

But, of course, Eq. (F18) contains the example presented in this work if we take  $\Lambda_{nm}(t) \equiv \lambda(t) \delta_{nm}$ . Equation (F18) motivates the natural generalization of a gauge transformation proposed in Eq. (20).

- 
- [1] B. Derrida, S. C. Manrubia, and D. H. Zanette, *Phys. Rev. Lett.* **82**, 1987 (1999).
- [2] B. Derrida, S. C. Manrubia, and D. H. Zanette, *Physica A* **281**, 1 (2000).
- [3] B. Derrida, S. C. Manrubia, and D. H. Zanette, *J. Theor. Biol.* **203**, 303 (2000).
- [4] S. Ohno, *Proc. Natl. Acad. Sci. USA* **93**, 15276 (1993).
- [5] J. H. Bennett, *Ann. Eugen.* **17**, 311 (1952).
- [6] J. H. Bennett, *J. Theor. Biol.* **4**, 28 (1963).
- [7] G. H. Hardy, *Sci. N. S.* **28**, 49 (1908).
- [8] W. Weinberg, Über den Nachweis der Vererbung beim Menschen. Jahresh. Ver. Vaterl. Naturkd. Württemb. **64**, 369 (1908) (English translations in Boyer 1963 and Jamenson 1977).
- [9] M. Begon, C. R. Townsend, and J. L. Harper, *Ecology—From Individuals to Ecosystems*, 4th ed. (Blackwell Publishing, Oxford, 2006).
- [10] A. T. Barucha-Reid, *Elements of the Theory of Markov Processes and Their Applications* (McGraw-Hill, New York, 1960).
- [11] M. Kijima, *Markov Processes for Stochastic Modeling* (Chapman & Hall, London, 1997).
- [12] M. Pagel, *Nature* **401**, 877 (1999).
- [13] K. L. Chung, *Markov Chains With Stationary Transition Probabilities*, 2nd ed. (Springer, New York, 1967).
- [14] L. Kleinrock, *Queueing Theory*, Vol. I (Wiley-Interscience Publication, New York, 1975).
- [15] M. Abramowitz and I. A. Stegun, *Handbook of Mathematical Functions* (Dover, London, 1972).
- [16] R. Feynman, *Phys. Rev.* **80**, 440 (1950).
- [17] W. Feller, *An Introduction to Probability Theory and Its Applications*, Vols. 1 and 2 (John Wiley & Sons Inc., New York, 1968).
- [18] M. Serva and L. Peliti, *J. Phys. A: Math. Gen.* **24**, L705 (1991).
- [19] E. Liberman, C. Hauert, and M. A. Nowak, *Nature* **433**, 312 (2005).
- [20] W. Heitler, *The Quantum Theory of Radiation*, 3rd ed. (Oxford University Press, London, 1954).
- [21] D. L. T. Rohde, S. Olson, and J. T. Chang, *Nature* **431**, 562 (2004).
- [22] N. T. J. Bailey, *J. Roy. Statist. Soc. Ser. B* **16**, 288 (1954).
- [23] D. G. Champernowne, *J. Roy. Statist. Soc. Ser. B* **18**, 125 (1956).
- [24] A. B. Clark, *The Time-Dependent Waiting Line Problem*, report M729-IR39 (University of Michigan, Ann Arbor, 1953).
- [25] W. Ledermann and G. E. H. Reuter, *Philos. Trans. Roy. Soc. London Ser. A* **246**, 321 (1954).
- [26] T. M. Apostol, *Mathematical Analysis* (Addison-Wesley Publishing Co., London, 1974).