Stochastic model for the species abundance problem in an ecological community

Simone Pigolotti,[1,2] Alessandro Flammini,[3,2] and Amos Maritan[3,4]

[1]*International School for Advanced Studies (SISSA-ISAS), Trieste, Italy*
[2]*Istituto Nazionale per la Fisica della Materia (INFM), UR-Trieste, Trieste, Italy*
[3]*Dipartimento di Fisica, Università di Padova, v. Marzolo 8, Padova, Italy*
[4]*Istituto Nazionale per la Fisica della Materia (INFM), UR-Padova, Padova, Italy*

We propose a model based on coupled multiplicative stochastic processes to understand the dynamics of competing species in an ecosystem. This process can be conveniently described by a Fokker-Planck equation. We provide an analytical expression for the marginalized stationary distribution. Our solution is found in excellent agreement with numerical simulations and compares rather well with observational data from tropical forests.

I. INTRODUCTION

One of the most widespread quantities employed in ecology to describe the biodiversity in a given ecosystem is the distribution of species abundance. In operational terms it can be defined as the histogram of the number of species (in a well-defined temporal and geographical context) consisting of a generic number of individuals, or, from a more theoretical perspective, as the probability that a generic species is composed by a certain number of individuals. Data collected in different locations suggests that the relative species abundance distributions show a certain degree of similarity [1]. To elucidate the causes that determine the shapes of these distributions and therefore their similarity is a problem of the utmost importance and not only of theoretical nature: to understand the motives that influences the relative rarity or commonness of different species can be of great help in determining policies for the conservation of the endangered ones.

The first studies on this subject can be dated back to the 1940s and are due to Fisher [2] and Preston [3]. Their works were focused on finding distributions that could fit well a particular data set in an empirical way. In particular, Preston [3] argued that the probability of finding species with a certain number of of individuals x should be lognormal distributed, while Fisher [2] proposed a function of the form e^{-ax}/x, with $a \ll 1$, the so-called Fisher log series.

Later, MacArthur [1] pointed out that similar distributions are found in very different ecosystems, suggesting that the shape of such distributions is to a large extent determined by very basic, general, and ecosystem-independent mechanisms. This in turn hinted at the possibility of predicting the shape of such distributions with simple and general models, without taking into account too many specific details of the ecosystem under consideration. Several models have been proposed that spoused this view [4–6]. Many of them restrict to modeling a single ecological community, a collection of similar species that feed on the same pool of resources in a local area. This definition implies that species belonging to the same community interact mainly in a competitive way: in particular, there are no prey-predator relationships among them. The particular case of a single ecological community can be framed in the wider context of a neutrality hypothesis. The concept of neutrality was introduced in the framework of a biomolecular evolution theory by Kimura [7], and then extended to other fields of biology. In the words of Hubbell [4], an ecological theory can be considered neutral when "…treats organisms in the community as essentially identical in their per capita probabilities of giving birth, dying, migrating and speciating. This neutrality is defined at the individual level, not the species level… ."

The question whether there exist ecological communities satisfying this assumption is still rather controversial [8], therefore it is crucial to understand what the consequences are of this zero-order hypothesis [9]. In the context of a neutral hypothesis, it is reasonable to describe the number of offspring to which any given individual gives place to as a stochastic variable. As a consequence, the number of individuals in a species at a given time can be regarded as a multiplicative random process.

Here we present a model aimed at reproducing the features of species abundance distributions under a minimal set of assumptions: neutrality and the possibility to describe the birth process as a multiplicative random processes. The model translates in a Fokker-Planck equation for the species abundance distribution and is amenable to an analytical treatment. The solutions found are compared with the experimental data that we have. The shape of these solutions depend on one parameter and give, in the two limiting cases, both a lognormal-like curve and the Fisher log series. The paper is organized as follows. In the second section we will present the model and comment on the assumptions made. In the third, we will take the continuum time limit of our model and will provide an analytical solution for the marginalized stationary probability distribution function (PDF). In the fourth section we will analyze what kind of connection there is between our model and the neutral theory [4,6]. In the last two sections we compare our results with the experimental data and comment on them.

II. THE MODEL

Let us consider an ecological community consisting of a fixed number s of species. According to the MacArthur and

Wilson theory of island biogeography [10], the number of species in a community approaches a dynamical equilibrium between immigration, speciation, and extinction. We assume that we can neglect the fluctuations around this equilibrium value; in our model, when a species become extinct, it is immediately replaced by another one. We also assume that the net effect of the competitive interaction between species in the community is just to keep the total number of individuals in the community fixed—the resources available are enough to support just N individuals across all the species. This last assumption implies that the populations of the species undergo a zero-sum dynamics. This hypothesis is well confirmed by experimental data [3,10]; at the end of Sec. III we will show that relaxing these constraints does lead to similar conclusions in the large N limit. We introduce the s variables x_i^t, representing the population of the ith specie at (discrete) time t, with the condition:

$$\sum_{i=1}^{s} x_i^t = N \quad \forall \, t.$$

Let $P(\lambda)$ be the probability that an individual in the community has λ offspring during one time step. Here neutrality plays a key role: our assumption implies that $P(\lambda)$ is the same for all individuals. The population of the ith species evolves according to the following equation:

$$x_i^{t+1} = N \frac{\sum_{k=1}^{[x_i^t]} \lambda_{k,i}^t + b}{\sum_{j=1}^{s} \left(\sum_{k=1}^{[x_i^t]} \lambda_{k,i}^t + b \right)} \tag{1}$$

where [ ] means the integer part. We are assuming that the existence of species with a noninteger number of individuals is not too drastic. This might lead to roundoff problems only for rare species. At each time step (generation) we just sum the number of offspring of every individual belonging to that species and then add a small quantity b. This quantity becomes relevant only for small x_i, and this describes the behavior of species near their extinction threshold. We are assuming that the net effect of extinctions, immigration, and speciation can be modeled in a simple way with this term, whose effect is to force the x_i's to be greater than zero. Indeed, for $b=0$, our system admits an absorbing state with only one x_i equal to N and the others equal to 0, the so-called monodominance [4]. Notice that species are only coupled through the denominator, which simply preserves the normalization condition.

The number of individuals of each species will be typically large, so we apply the central limit theorem to the sum of random variables in this equation, obtaining the following model:

$$x_i^{t+1} = N \frac{\bar{\lambda} x_i^t + \sigma \sqrt{x_i^t} \xi_i^t + b}{\sum_{j=1}^{s} (\bar{\lambda} x_j^t + \sigma \sqrt{x_j^t} \xi_j^t + b)} \tag{2}$$

where $\bar{\lambda}$ and σ are the mean value and the root mean square deviation of the distribution $P(\lambda)$, and the ξ's are uncorrelated Gaussian variables with zero-mean and unit variance.

It is worth noting the relation between our model and the multiplicative process introduced by Kesten in [11]. Kesten studied random multiplicative processes of the form $X_{t+1} = \lambda_t X_t + b_t$, where X_t is the variable and both λ and b are random variables. He found that, depending on the mean value of λ and on the boundary conditions, one retrieves a lognormal or a power-law regime. Models for ecology and economics based on this kind of processes were proposed by Sornette [12] and Solomon [13]. In our model the number of individuals of different species can be thought as following coupled Kesten-like processes. The coupling is a consequence of the constrain that keeps fixed to N the number of individuals in the community and that is enforced in Eq. (1) by the factor N and by the denominator.

III. THE CONTINUUM LIMIT

In order to obtain some analytical result, we take the continuous time limit of this model, by introducing the time interval dt in the following way:

$$\lambda \rightarrow 1 + \lambda dt$$

$$b \rightarrow b dt$$

$$\sigma \rightarrow \sigma dt. \tag{3}$$

By means of this substitution, our model becomes

$$x_i^{t+dt} = \frac{x_i^t + dt(\bar{\lambda} x_i^t + \sigma \sqrt{x_i^t} \xi_i^t + b)}{1 + \frac{dt}{N} \sum_{j=1}^{s} (\bar{\lambda} x_j^t + \sigma \sqrt{x_j^t} \xi_j^t + b)}. \tag{4}$$

Expanding the denominator and using the fact that $\sum_j x_j = N$, we get the Langevin equation

$$\dot{x}_i = f_i(x) + \sum_{j=1}^{s} B_{ij}(x) \xi_j \tag{5}$$

where

$$f_i(x_i) = b \left(1 - \frac{s}{N} x_i \right)$$

$$B_{ij}(\underline{x}) = \left(\delta_{ij} - \frac{x_i}{N} \right) \sqrt{x_j} \, . \tag{6}$$

The Fokker-Planck equation [14] associated with this Langevin equation is

$$\dot{P}(\underline{x},t) = -\sum_{i=1}^{s} \partial_i \left\{ -f_i P(\underline{x},t) + D \sum_j \partial_j [g_{ji}(\underline{x}) P(\underline{x},t)] \right\} \tag{7}$$

with $D = \sigma^2 / 2$ and

$$g_{ij}(\underline{x}) = g_{ji}(\underline{x}) = \sum_k B_{ik} B_{jk} = \left(\delta_{ij} - \frac{x_j}{N} \right) x_i. \tag{8}$$

We search for a solution of this equation satisfying detailed balance, i.e., $P^{st} f_i = D \sum_j \partial_j (g_{ij} P^{st})$. Defining the marginalized PDF
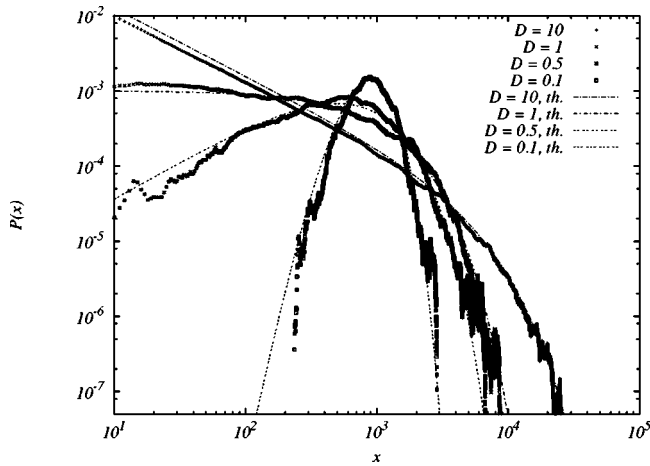
FIG. 1. Simulation of marginalized stationary pdf for various values of the diffusion coefficient D, compared with theoretical curves. For all curves $b=1$, $\bar{\lambda}=1$ $s=100$, $N=10^9$. Curves are binned linearly with binning size $\delta x=10^4$. Notice that as D increases the curve approaches the Fisher log series.

$$p(x) = \int_0^\infty \prod_{j \neq i} dx_j P^{st}(\underline{x}) \qquad (9)$$

we can easily obtain an equation for $p(x)$

$$b\left(1 - \frac{sx}{N}\right)p(x) = D\frac{d}{dx}\left[\left(x - \frac{x^2}{N}\right)p(x)\right]. \qquad (10)$$

This equation can be easily solved, giving:

$$p(x) \propto x^{\beta-1}\left(1 - \frac{x}{N}\right)^{\beta(s-1/N)-1} \qquad \beta = \frac{b}{D} \ . \qquad (11)$$

Notice that this distribution correctly shows the mono-dominance behavior $\delta(0)$ or $\delta(N)$ in the limit $\beta \to 0$. Finally, if we fix $\mu = \beta s/N$, in the limit for $N \to \infty$ we obtain

$$p(x) = \frac{\mu^\beta}{\Gamma(\beta)x^{1-\beta}}e^{-\mu x}. \qquad (12)$$

In Fig. 1 we plot simulation of the stationary PDF for various value of the parameter β, and check the validity of (12).

Instead of having a system of stochastic differential equations, it is possible to take into account the interaction of a species with the ecosystem in an averaged way. Let us consider the Langevin equation:

$$\dot{x}(t) = b + \bar{\lambda}x - \gamma x + D\sqrt{x}\xi \qquad (13)$$

where the parameter γ takes into account the effect of competition. In order to have normalizable solutions, we have to require that $\gamma > \bar{\lambda}$. When this condition holds, it is straightforward to show that the stationary pdf satisfying detailed balance is the same as (12), with $\mu = -(\bar{\lambda} - \gamma)/D$. Notice that in this case, the detailed balance solution is exact; it is also remarkable that the stationary distribution (12) can be achieved without fixing neither the number of species, nor the number of individuals.
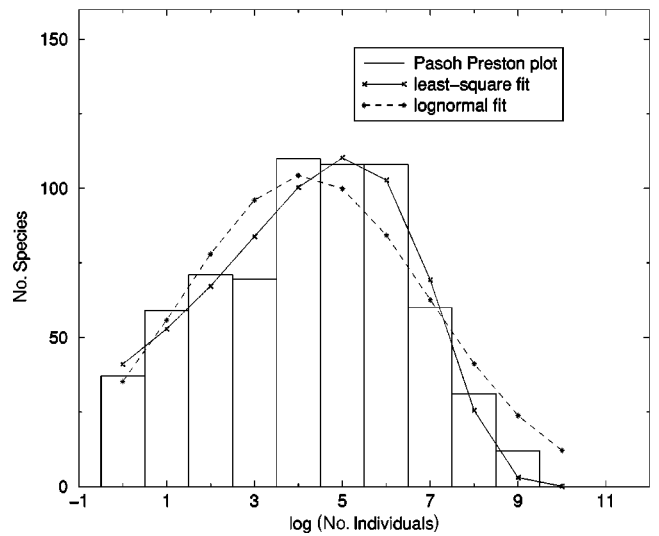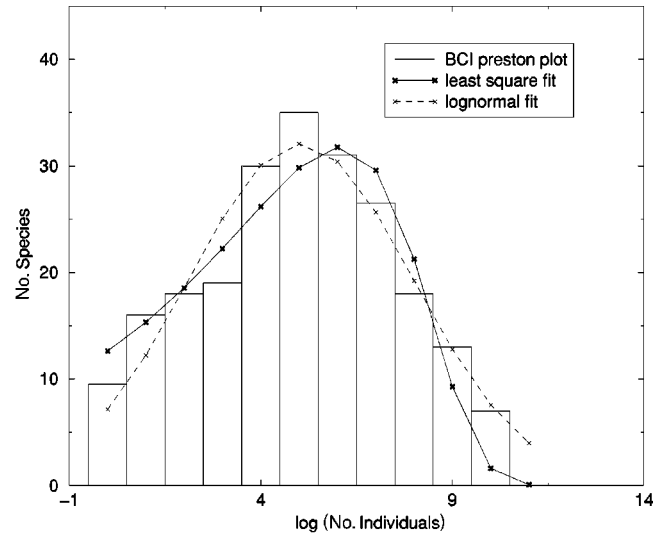

FIG. 2. Fit of Barro Colorado Island (BCI) and Pasoh species abundance data, Preston plot [4]. Comparison between our solution and lognormal. Fitted value of the parameters of our distribution are $\beta=0.23$ and $\mu=0.010$ for the BCI; $\beta=.37$ and $\mu=0.015$ for Pasoh. In absence of an objective estimate of the error bars on the observational data, both our result and the lognormal give a reasonable fit.

IV. CONNECTION WITH THE NEUTRAL THEORY

An interesting question is whether there is some relationship between our model and the neutral theory [4], as formulated by Volkov *et al.* [6] (see also [5]). More precisely, one could ask if our model arises from the continuum limit of a master equation similar to the one proposed for the neutral theory. Let us write the master equation for the generic birth and death process

$$\dot{P}(x) = d(x+1)P(x+1) + b(x-1)P(x-1)$$
$$- [d(x) + b(x)]P(x) \qquad (14)$$

for $x > 1$. We set $d(1)=0$ to avoid that a species disappears without being replaced by another one [6]. Equation (14) can

be converted into a Fokker-Planck equation assuming that x is a continuous variable and that $b(x)$ and $d(x)$ and $P(x)$ are smooth enough that we can expand them into a Taylor series. Thus, for example,

$$d(x+1)P(x+1) - d(x)P(x) = \frac{\partial}{\partial x}[d(x)P(x)]$$
$$+ \frac{1}{2}\frac{\partial^2}{\partial x^2}[d(x)P(x)].$$

Taking the first-order expansion, we obtain from (14) the following Fokker-Planck equation:

$$\dot{P}(x) = -\frac{\partial}{\partial x}J(x) \qquad (15)$$

with

$$-J(x) = [d(x) - b(x)]P(x) + \frac{1}{2}\frac{\partial}{\partial x}\{[d(x) + b(x)]P(x)\} + \dots .$$

$J(x)$ has the meaning of a probability current and we can write the general form of the stationary solution satisfying detailed balance as a function of $d(x)$ and $b(x)$ setting $J(x) = 0$

$$p(x) = \frac{p(1)}{d(x) + p(x)}\exp\left[-\int_1^x \frac{d(x') - b(x')}{d(x') + b(x')}dx'\right]. \quad (16)$$

One can easily check that our stationary PDF (12) is recovered, provided the following particular choice of the birth and death coefficients:

$$d(x) = \frac{1+\mu}{2}x - b$$
$$b(x) = \frac{1-\mu}{2}x + b. \qquad (17)$$

This choice [6] implies that there is balance between immigration and emigration in each species. The more general case, in which this equilibrium does not hold, is treated in [15].

V. COMPARISON WITH EXPERIMENTAL DATA

Among the most reliable data on single-trophic species distribution of species abundance is the tropical forest census [16]. In order to make a coarse-graining, a Preston plot is used; data are collected via a logarithmic binning in base 2, and species at the edge between two consecutive binning are equally divided between them. Since we have a continuous probability density, we compared the histogram with the integral over the bins of the distribution with the experimental data and made a least-square fit of the parameters β and μ, plus the normalization. We found a good agreement of our predicted curve with the histogram; in Fig. 2 the comparison between our solution and the lognormal is shown. Notice that the two distributions have the same number of fitted parameter. It would be interesting to compare our distribution with data collected form other kind of ecosystems and to try to clarify the dependence of our free parameter β from ecological quantities, such as the immigration pressure, the speciation rate, and the extinction threshold.

VI. DISCUSSION AND PERSPECTIVES

The model we introduce admits a family of stationary pdf depending on the parameter β. This parameter fully determines the shape of the distribution: for $\beta \ll 1$, one recovers the Fisher log series; while for β large, one obtains a lognormal-like distribution. As we already pointed out, both of these distributions are well known in the population biology literature as possible candidates to be the "right" distributions found in nature.

There is some analogy between our model and the Kesten process. Indeed, also the Kesten process admits two different regimes: one lognormal and one with a power-law tail. The main differences is that in our case the multiplicative random process is applied to the square root of the variables rather than to the variable itself. As a consequence, in the Kesten case, the exponent of the power-law tail of the stationary distribution is always greater than one, while the small β regime of our system is characterized by a power-law tail over many decades, with an exponent that is always less than 1. The cutoff due to the conserved number of individuals ensures the normalization of these long-tailed distributions.

It is remarkable that our distribution is the same as found in studies made by Kerner in the 1950s [17] on the invariant measure in a system of Lotka-Volterra equations with purely asymmetric couplings. In that work, the interactions are only of predator-prey type, and the system is deterministic, while we are considering a stochastic system with purely competitive coupling. The discovery of the same distribution in such different models suggests that there might exist some deeper and more general mechanism determining the statistical behavior of the ecosystems, regardless of the type of interactions among species.

[1] R. MacArthur, Proc. Natl. Acad. Sci. U.S.A. **43**, 293 (1957).

[2] R. A. Fisher, A. S. Corbet, and C. B. Williams, J. Anim. Ecol. **12**, 42 (1943).

[3] F. W. Preston, Ecology **29**, 254 (1948)

[4] S. P. Hubbell, *The Unified Neutral Theory of Biodiversity and Biogeography* (Princeton University Press, Princeton, NJ, 2001).

[5] A. McKane, D. Alonso, and R. V. Solé, Phys. Rev. E **62**, 8466 (2000); e-print physics/0305022.

[6] I. Volkov, J. R. Banavar, S. P. Hubbel, and A. Maritan, Nature (London) **424**, 1035 (2003).

[7] M. Kimura, *The Neutral Theory of Molecular Evolution* (Cam-

bridge, University Press, Cambridge, UK, 1983).

[8] D. Tilman, *Plant Strategies and the Dynamics and Structure of Plant Communities* (Princeton University Press, Princeton, NJ, 1988).

[9] J. Harte, Nature (London) **424**, 1006 (2003).

[10] R. MacArthur and E. O. Wilson, *The Theory of Island Biogeography* (Princeton University Press, Princeton, NJ, 1967).

[11] H. Kesten, Acta Math. **131**, 207 (1973).

[12] D. Sornette, Physica A **250**, 295 (1998).

[13] S. Solomon, e-print cond-mat/9901250.

[14] C. W. Gardiner, *Handbook of Stochastic Methods* (Springer-Verlag, Berlin, 1985).

[15] I. Volkov, J. R. Banavar, S. P. Hubbel, and A. Maritan, pre-print.

[16] R. Condit, S. P. Hubbell, and R. B. Foster, J. Trop. Ecol. **12**, 231 (1996).

[17] E. H. Kerner, Bull. Math. Biophys. **19**, 121 (1957); Bull. Math. Biophys. **21**, 217 (1959).