

Exercise 03 - Neutral theory and emergent patterns in Ecology

Federico Agostini, Federico Bottaro, Gianmarco Pompeo

1 Introduction

We are once again presented with the dataset containing information on a 2005 tree census performed on a 50-hectare piece of land in the Barro Colorado Island in Panama.

The purpose of this project is to apply Neutral theory to a well-mixed ecosystem with a high biodiversity, which we will analyze using the Voter model; as a variation, we will also build a logistic growth model for the population dynamics in the very same system.

2 Preliminary operations

To begin with, we select only the species which are classified as alive in the dataset. We then divide the overall piece of land into $N = 800$ equally-sized subplots which we will assume to be independent in order to be able to perform statistical analysis on them.

For each subplot, we count the abundance for each species, that is the absolute number of individuals present in there. We will obtain a matrix \mathbf{A} where each row corresponds to a subplot and each column to a species; this will therefore be an $N \times S$ matrix, where S is the total number of species in the island.

After doing so, we reshape this matrix into one big $N \times S$ vector, but we will trim it right away to remove all the 0 components. We obtain a new vector \mathbf{X} which contains all information about present species in our sample, but we clearly lose the structure of the previously defined matrix, meaning we do not have any order on the species nor on the corresponding subplots.

At this point, we build the empirical cumulative distribution function of the Relative Species Abundance (RSA); $P_>(x)$ will give us the probability that by picking at random any species, its abundances will be greater or equal than x , with the initial obvious condition that $P_>(1) = 1$.

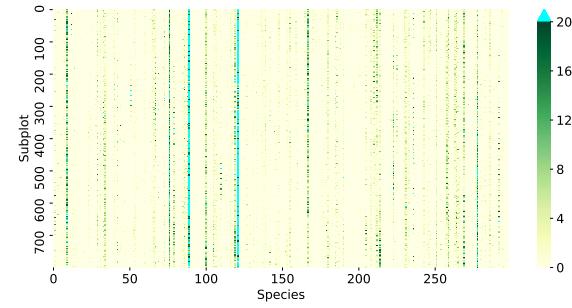


Figure 1: Heatmap of the abundance matrix \mathbf{A} .

3 Comparison with the Master Equation - stationary regime

The general birth-death Master Equation (ME) found in the Voter model applied to Ecology is

$$\frac{dP_n}{dt} = b_{n-1}P_{n-1}(t) + d_{n+1}P_{n+1}(t) - (b_n + d_n)P_n(t) \quad (1)$$

where P is the probability that randomly choosing a species, this has n individuals at a given time and the coefficients are the birth rate and death rate, respectively

$$b_n = bn \quad d_n = dn \quad (2)$$

In particular, the value b_0 is often referred to as immigration rate and it is experimentally known to be

$$b_0 = m = 0.05 \quad (3)$$

The stationary solution of the ME, obtained by imposing the time derivative to be null, can be written as

$$P_n = NP_0\nu(1-\nu)^{n-1}\frac{1}{n} \quad (4)$$

with ν another label for the migration rate. If we now impose normalization by requiring $\sum_{n=1}^{\infty} P_n = 1$ we get

$$P_n^* = -\frac{(1-\nu)^n}{\log \nu} \frac{1}{n} \quad (5)$$

which is also known as *Fisher log-series* distribution.

We wish to compare the cumulative function obtained from this stationary solution with the cumulative of the RSA. Using the empirical cumulative distribution function from **X** obtained in Section 2 we can plot both the experimental and the theoretical curve; in particular we also show the results in a logarithmic scale to emphasize the differences between the two curves.

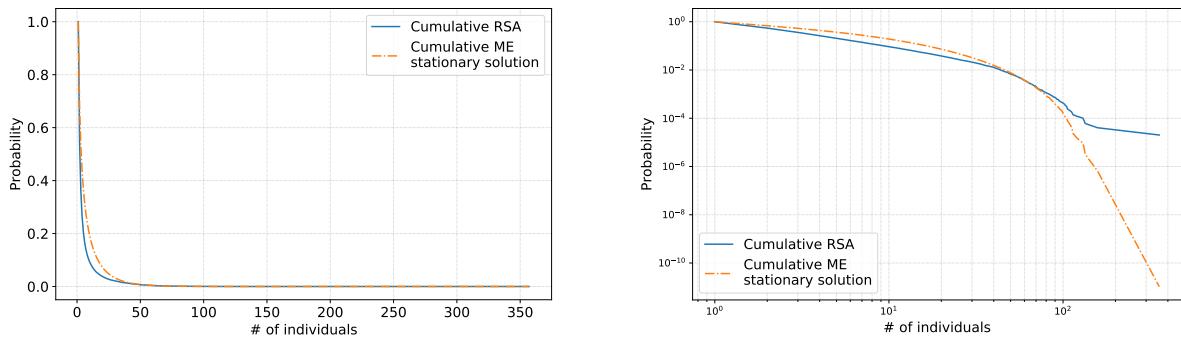


Figure 2: Comparison between the cumulative Relative Species Abundance and the cumulative distribution of the stationary solution of the Master Equation. The graph on the right-hand side is in log-log scale.

As shown in Figure 2 the data follows the theoretical prediction quite closely except for a tail due to the fact that there is a species which is dominant and shows an exceptionally high abundance, which appears as an outlier in the log-scale.

4 Empirical species-area relationship

In this section we try to build the empirical species-area relationship (SAR) for our ecosystem. To do so, we want to investigate the number of different species as a function of the area covered by the sample under consideration.

Operatively, we divide the dataset in decreasing number of subplots, from 800 to 1. Since no matter the number they can still be considered as independent samples, for each division we take the mean value of the number of different species found in each subplot; the error on this estimate is derived as the error on the mean.

Fitting for the parameter m , we plot the number of species as a function of the area $s(a)$ and compare it with the prediction from Neutral Theory in a well-mixed ecosystem, given by

$$\langle s(a) \rangle = S \left(1 - \frac{\log(\frac{a}{A}(1-m) + m)}{\log(m)} \right) \quad (6)$$

where S is the total number of species, a is the area of the considered subplot and A is the total area of the Barro Colorado forest.

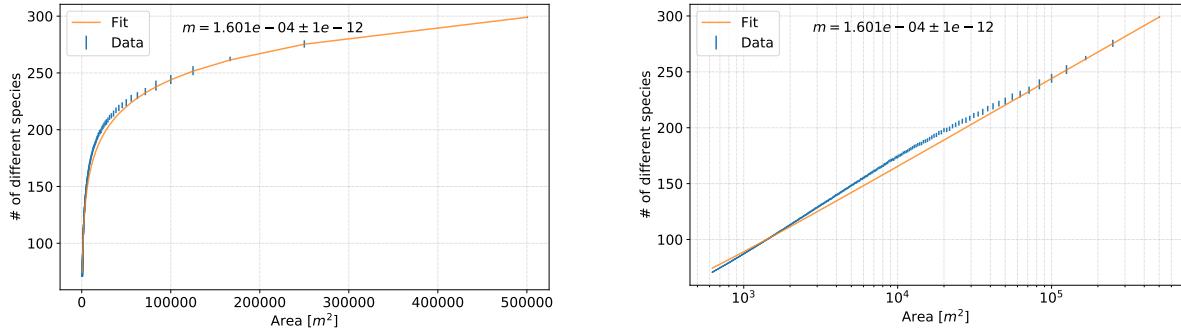


Figure 3: Empirical Species Area Relationship: the fit is done using Equation 6, where the parameter m is retrieved from the data. The graph on the right-hand side is in logarithmic x-axis scale.

In Figure 3 it is possible to visualize the result with the relative error bars. As we can see the error increases as the area of the subplots grows due to the fact that we have fewer samples available for computing the mean value as we increase the size of the subplots.

The fit also provides with an estimation for the parameter m

$$m = 1.601\ldots \cdot 10^{-4} \pm 1 \cdot 10^{-12}$$

which has an impressively low value for the error considering that this fitted value of migration rate is completely off the given estimate of $m = 0.05$. The most reasonable conclusion is that through the Species-Area Relationship model we are completely incapable of modelling our ecosystem.

In the log-scaled plot it is possible to appreciate how the data are distributed in a fairly linear fashion, except for a deviation for the middle-sized subplots.

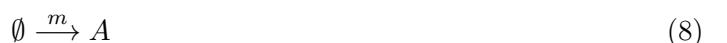
5 Logistic growth

Logistic growth is also effective in describing the dynamics of an ecosystem. The idea is that species can grow exponentially until the system allows it, meaning that there will be a limit after which the growth saturates: this limit is called *carrying capacity* \mathcal{K} of the system.

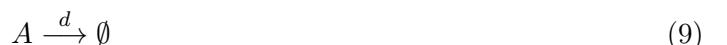
We consider the simplified case where only one species A is present and we define the possible reactions that it can undergo. A birth reaction is of the kind



where b is the birth rate per unit of time and individual. However, we also consider the case of migration, so that there is a possibility for A to grow by one unit according to the reaction



even if no birth takes place or there are no A individuals at all; m is of course the migration rate as a function of time. Lastly, the death reaction can be written as



where d is the death rate per time and particle and which we set equal to

$$d = \frac{bA}{\mathcal{K}} \quad (10)$$

By this way, we are able to incorporate the information on the carrying capacity of the system in the death reaction, which will prevent an explosion above the system limit apart from fluctuations. We can now compute the transition rates; one derives from the first two reactions,

$$\mathcal{W}^+(A) \equiv \mathcal{W}(A \rightarrow A + 1) = bA + m \quad (11)$$

while from the death reaction we retrieve for the other one

$$\mathcal{W}^-(A) \equiv \mathcal{W}(A \rightarrow A - 1) = Ad = \frac{bA^2}{\mathcal{K}} \quad (12)$$

We hence write the corresponding ME in the following form

$$\frac{dP_n}{dt} = P_{n-1}(bA + m) + P_{n+1} \frac{bA^2}{\mathcal{K}} \quad (13)$$

with the same formalism as in Section 3; by setting it equal to 0 we proceed on to calculating the stationary state, which after some mathematical manipulations is found to be

$$P_n^* = \prod_{n=1}^A \frac{b(n-1) + m}{\frac{bn^2}{\mathcal{K}}} = \frac{\prod_{n=0}^{A-1} bn + m}{(\frac{b}{\mathcal{K}})^A (A!)^2} = \frac{[(A-1) + \frac{m}{b}][(A-2) + \frac{m}{b}] \dots [\frac{m}{b}]}{\frac{b}{\mathcal{K}^A} (A!)^2} \quad (14)$$

with the proper normalization.

To see the results in a concrete case, we set

$$m = 0.1 \quad b = 1 \quad \mathcal{K} = 10$$

and plot the probability of having a given number of individuals for the species A as a function of the number of individuals itself, considering the stationary state.

We also wish to include a simulation in order to study the agreement with the theoretical results. We initialize at random the number of starting individuals, which can undergo the reactions listed above with given probabilities according to the implementation of the Gillespie algorithm; we ran the algorithm for 10^8 iterations.

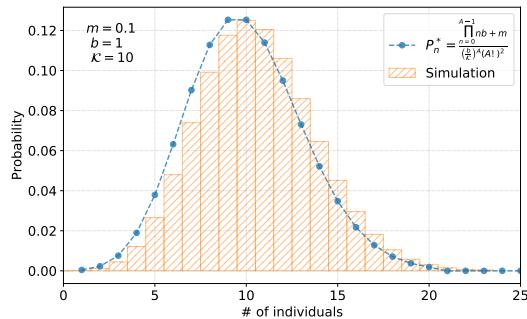


Figure 4: Probability as a function of the number of individuals for the logistic growth model. The histogram refers to the simulation.

In the plot, we can see that the probability distribution has a peak around \mathcal{K} and it decreases afterwards, which is expected. This is coherent with the model we are trying to develop, since we assume our system has a limited number of individuals it can support with its resources. If we happen to have more individuals than this value due to stochastic fluctuations, death reactions will occur so as to bring it back below the carrying capacity.

The comparison with the simulation is satisfactory, as also the histogram obtained from the data is peaked around the carrying capacity of the system.