

Exercise 3

Neutral Theory and emergent pattern in Ecology

Paolo Frazzetto, Enrico Lorenzetti

Università degli Studi di Padova
Statistical Mechanics of Complex Systems
Prof. S. Suweis, Prof. A. Maritan

15 May, 2019

Introduction

In this assignment we analyse an application in Ecology of the Voter Model, that was originally developed to describe the dynamics of opinion spreading among individuals. The theoretical results have been compared with data coming from the *Barrio Colorado Forest Census Data*. All the calculations and plots have been carried out by means of the attached `Python` notebook and `Mathematica`.

1 Data Pre-processing

The original 50-hectares has been split in $N = 800$ subplots of equal area, as shown in Figure 1. By doing so, we assume that these subplots are independent and we can compute statistics on them. Then we calculated the vector of the abundances for all the species (x_1, \dots, x_S) for each subplot. In

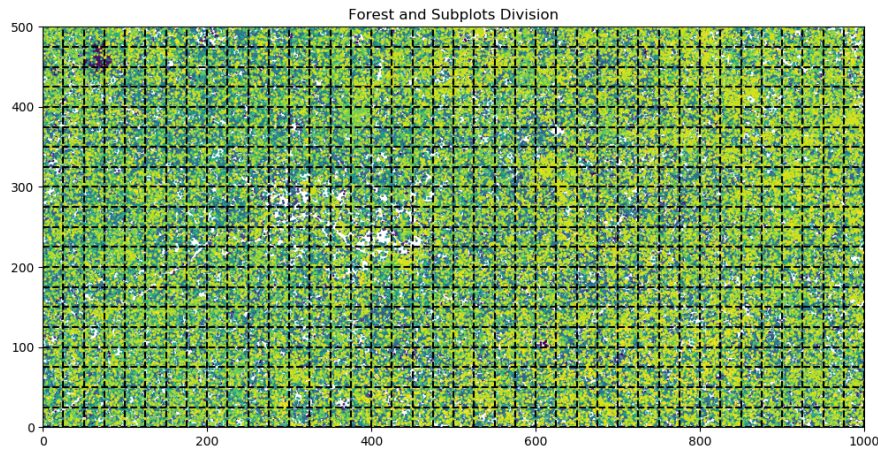


Figure 1: Division of Barrio Colorado Forest in 800 subsets.

this way we obtain a $N \times S$ matrix where each row represents the statistics of abundances in one subplot.

2 Relative Species Abundance

We are interested to study the *Relative species abundance (RSA)*, i.e. the probability that a species has n individuals in a given subplot. If we count the number of occurrences of the data matrix, it is simple to build the RSA cumulative distribution. We define $P_{\geq}(n)$ as the probability that picking a species at random it has an abundance greater or equal than n , as it is shown in Figure 2. Obviously it holds that $P_{\geq}(1) = 1$. This plot is limited at $n = 120$ individuals since there is an outlier subplot with 328 individuals of a certain species that leads to a extremely low-but-not-null probability. The

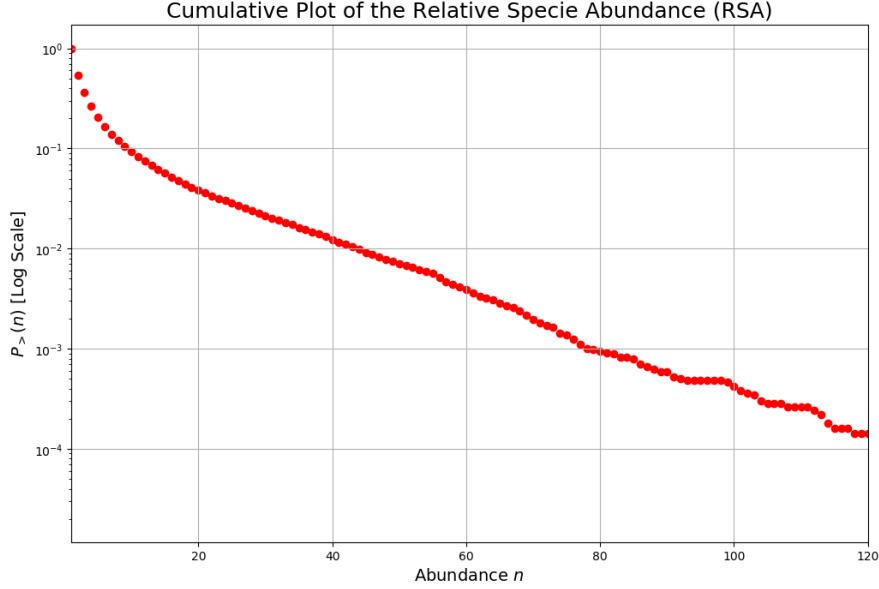


Figure 2: RSA Cumulative Distribution obtained from data.

distribution decreases steeply (notice the log scale of the vertical axis) meaning that there are few species whose individuals are very common and numerous, while for most of them they are scarcer.

3 Neutral theory

To explain the shape of Figure 2, we conceive a model in which each species is independent from the others and the stochastic system is ruled by the *Voter Model*. In this way, we can write the master equation, whose stationary solution shows us the probability distribution of the abundance of a single species. Exploiting the fact that for hypothesis species are independent, we need to perform these counts just for one species. Taking into all these facts, we arrive at the following equation:

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

Where b_n is the transition rate $W(n \rightarrow n+1)$, while $d_n = W(n \rightarrow n-1)$, with the physical boundary condition that

$$\begin{cases} d_0 = 0 \\ b_{-1} = 0 \\ b_0 = m \end{cases}$$

In order to find P_n^* , the stationary solution, we impose that $\left. \frac{dP_n}{dt} \right|_{P_n^*} = 0$. If we define $J(n) = d_n P_n - b_{n-1} P_{n-1}$, equation 1, with the notation $P_n^* \equiv P_n$, gets

$$J(n+1) - J(n) = 0 \quad (2)$$

Therefore

$$\sum_{x=0}^n J(x+1) - J(x) = 0 \quad (3)$$

It's easy to see that this condition is equivalent to

$$J(n+1) = J(0) = 0 \quad (4)$$

So

$$P_n = \frac{b_{n-1}}{d_n} P_{n-1} = \frac{b_{-1}}{d_n} \frac{b_{n-2}}{d_{n-1}} \dots \frac{b(0)}{d(1)} P_0 \quad (5)$$

As we know for Voter Model in big ecosystem, i.e. neglecting $\mathcal{O}(\frac{n}{N})$ higher orders, the transition rates are

$$\begin{cases} b(n) \approx (1-m)\frac{n}{N} \\ d(n) \approx \frac{n}{N} \end{cases}$$

and equation 5 becomes

$$P_n = \frac{(1-m)^{n-1} (\frac{1}{N})^{n-1} (n-1)! m}{(\frac{1}{N})^n n!} P_0 \quad (6)$$

Therefore

$$P_n = \frac{N}{n} m (1-m)^{n-1} P_0 \quad (7)$$

With the condition of normalisation

$$\sum_{n=1}^{\infty} P_n = 1 \quad (8)$$

we finally arrive at the formula for RSA

$$P(n) = -\frac{(1-m)^n}{n \log(m)} \quad (9)$$

From that, we can simply obtain the the number of species with n trees multiplying the formula by the total number of species. If we want to obtain the cumulative function we just sum over n as follows

$$P_{CDF}(x) = \sum_{n=x}^{\infty} -\frac{(1-m)^n}{n \log(m)} \quad (10)$$

with the obvious property that $P_{CDF}(1) = 1$, as we would suppose. Figure 3 compares the theoretical values obtained with $P_{CDF}(x)$ and previous data. Even though they display a similar behaviour, they do not match. One possible explanation is that this tight area division favours lower abundance values.

4 Species-Area Relationship

In this section we would like to confront the prediction of *Neutral Theory* with our experimental data, concerning the mean number of different species found with the increasing of size of the sample area (SAR). In order to evaluate that, we simple divide the forest in 800 subplots as seen before, with the condition that the total area is equal to the sum of all of them, i.e. $A_{tot} = \sum_{k=1}^{800} a_k$. For construction, the numeration of the subplots preserves the space of the forest, since for example (a_1, a_2) are close

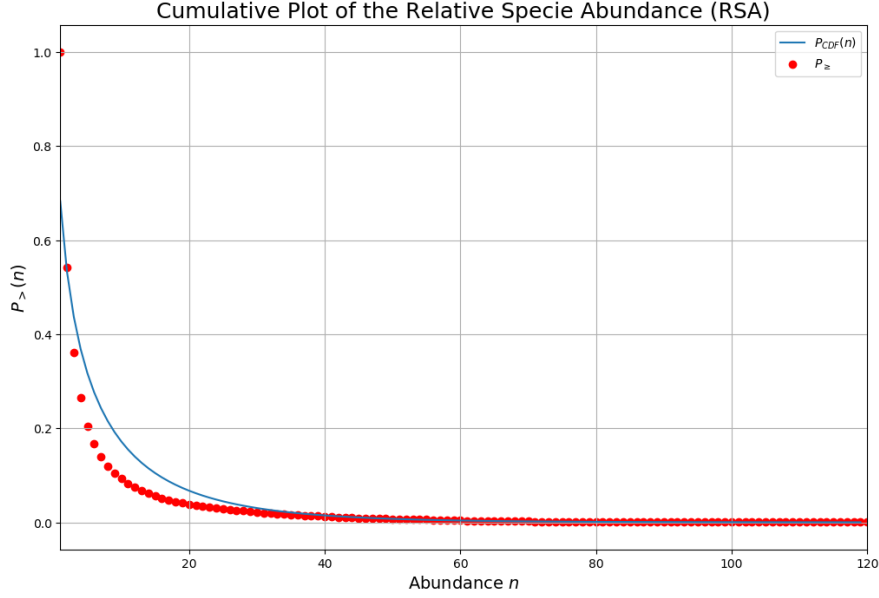


Figure 3: RSA Cumulative Distribution compared with theoretical expectations.

subplots, (a_2, a_3) too and so on. Nevertheless, we are supposing them as independent realisations, so to lose this information we prefer to shuffle the relative associated subplot number. This allows us to perform average among subplots which can be very distant from each other, therefore they more reasonably respect the hypothesis of independence. We plot the mean number of species found in a given relative area $\alpha = \frac{ja}{A_{tot}}$, with j from 1 to 800 and a the unit area of the subplot, obtained mediating over all possible non overlapping clusters built as $\cup_{k=k^*}^{k^*+j-1} a_k$. The results are shown in Figure 4.

From theory, if we resort to the hypothesis of *random placement*, which guarantees us that the disposition of the individuals is random and uniform inside the total area, we can write the probability of i -th species of belonging p_i and not belonging q_i to the α area as

$$\begin{cases} q_i = (1 - \alpha)^{n_i} \\ p_i = 1 - (1 - \alpha)^{n_i} \end{cases}$$

If we express the generating function as $G(z) = \prod_{i=1}^S (q_i + zp_i)$, we can obtain the mean number of different species expected in the area α

$$\bar{s}(n) = \left. \frac{d}{dx} \log(G(e^x)) \right|_{x=0} = S - \sum_{i=1}^S (1 - \alpha)^{n_i} \quad (11)$$

However, n_i in the *Voter Model* is a random variable, so we have to perform the convolution between the $\bar{s}(n)$ and $P(n) = \prod_i^S P(n_i)$ taken from the expression 9. Once again, $P(n)$ is simply the product of $P(n_i)$ because we are working with the Neutral Theory hypothesis. Finally we arrive at the formula for the *SAR*

$$\langle s(\alpha) \rangle = S \left(1 - \frac{\log(m(1 - \alpha) + \alpha)}{\log(m)} \right) \quad (12)$$

To get an estimation of m , we fit data using this formula and confront it with the value reported in the literature $m = 0.05$. We got that $m_{fit} = (1.49 \pm 0.08) \times 10^{-5}$. As previously noticed, this results remark the fact that the hypothesis of our model are not suitable for this system, or at least with this division of subplots.

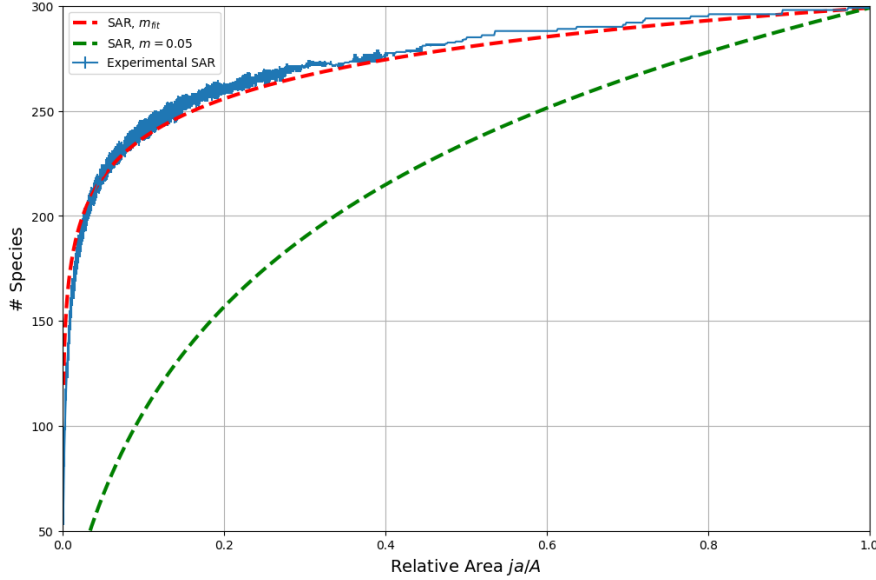


Figure 4: SAR for the Voter Model. Notice that error bars are present only in the first half of the plot, since after that there is only one possible sample of such size. Therefore, the m_{fit} has been computed without taking into account these errors.

5 Logistic Growth

In conclusion, we briefly study the *Logistic Growth* model, characterised by the carrying capacity K , which prevents unlimited growth. Indeed, this parameter takes into account the fact that the resources in the system are finite. The transition rates can be written as

$$\begin{cases} W(n \rightarrow n+1) = (nb + m) \\ W(n \rightarrow n-1) = \frac{n^2 b}{K} \end{cases}$$

So, using the recursive formula 5 previously obtained, with $b_n = W(n \rightarrow n+1)$ and $d_n = W(n \rightarrow n-1)$ we get

$$P_n = \frac{K^n (b(n-1) + m)(b(n-2) + m) \cdots (1+m)m}{b^n (n^2)!} P_0 \quad (13)$$

which we can rewrite as

$$P_n = \sum_{x=0}^{n-1} \frac{K^n b^{-1-x} m^{x+1} (n-1)!}{(n^2)! x!} P_0 \quad (14)$$

To find the correct distribution for $b = 1$, $m = 0.1$ and $K = 10$ we should normalise P_n imposing 8. Notice that once again we are considering only species alive. Nevertheless, for numerical reasons, we cannot perform a sum over all n but we have to put a cut-off. Considering the power of our computational machine and the fact that $\frac{P(90)}{P(K)} \approx 10^{-53}$, we set $n_{max} = 90$. As you can see, differently from the deterministic model, we can have $P_n \neq 0$ even when $n > K$. This is mainly due to the fluctuations of our stochastic process. Even so, it is obvious that when $n \gg K$, $d_n > b_n$ and the population tends to decrease the number of individuals rather than thriving. It is also reasonable that $x_{mean} \approx 8.98 \neq K$ because in the deterministic model the carrying capacity is an asymptotic limit, therefore never reached, so the noise tends to have its mean slightly below K . We finally run some simulations using the *Gillespie Algorithm*. The results show that, for any initial condition, the system stabilise around $n \approx K$.

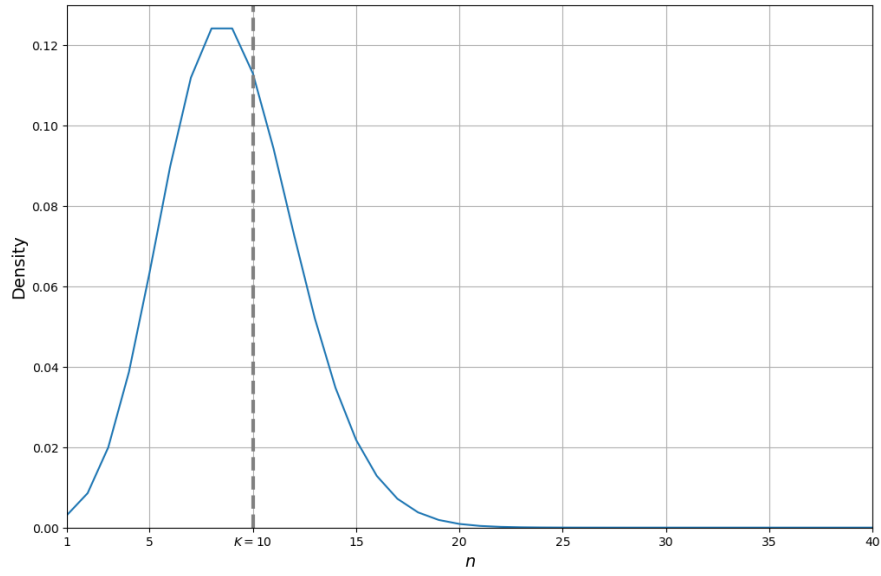


Figure 5: Probability distribution of the Logistic Growth Model considering only the alive species.

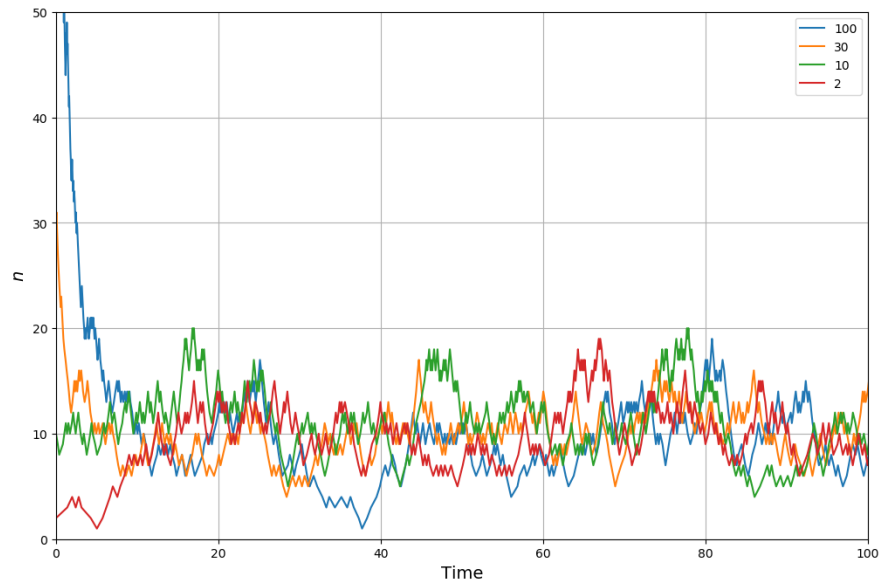


Figure 6: Gillespie simulation for different starting amounts of individuals.