

Università degli Studi di Padova

Dipartimento di Fisica e Astronomia “Galileo Galilei”

Corso di Laurea in Physics of Data

EX 3 STATISTICAL MECHANICS OF COMPLEX SYSTEM

OF

MARCO AGNOLON, TOMMASO TABARELLI

Anno accademico 2018-2019

The system

Our task was to analyze the trees distribution of the forest of the Barro Colorado Island. The database contains many information about every tree in the selected fifty hectares ($1000m \cdot 500m$ area), among them we considered only the positions and the species of the trees. The position are given by a couple of number which represent the two-dimensional coordinates of the tree. We set a number for each species to have a one to one identification of the species with a specific number up to 299, i.e. the total number of species.

Task 1 and Task 2

We divided the fifty hectares in 800 subplots of 0.0625 hectares each ($25m \cdot 25m$ subplots area). Then we mapped each tree in the correspondent subplot. Starting from this information, we then evaluated the abundance of each species in each subplot. We flattened it and used it as independent samples.

From this vector we build the empirical cumulative distribution function calculating the probability that picking a species at random, it has an abundance equal or greater than x , i.e. $P_{\geq}(x)$, with $P_{\geq}(1) = 1$. This is the cumulative distribution of the Relative Species Abundance (RSA).

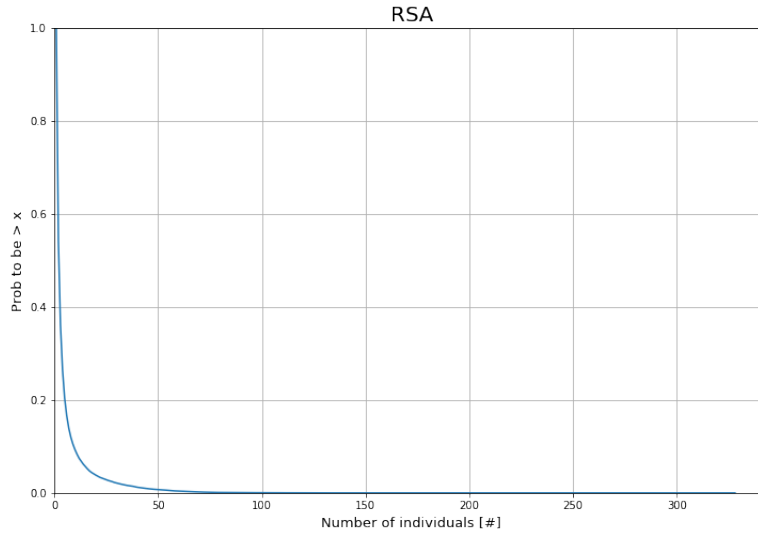


Figure 1: Relative species abundance of the flattened vector.

As expected, the RSA shows that the greater the number of individuals of a species in a subplot, the smaller the probability to find them together.

Task 3

We compared the empirical cumulative RSA of the Task 2 with the stationary solution of the birth and death Master Equation, which is:

$$\begin{aligned} \frac{d}{dt}P_n(t) &= b(n-1) \cdot P_{n-1}(t) + d(n+1) \cdot P_{n+1}(t) - \\ &\quad - (b(n) + d(n)) \cdot P_n(t) \end{aligned}$$

and its solution in the stationary case, $\frac{d}{dt}P_n(t) = 0$, is:

$$P_n^* = v \frac{(1-v)^n}{\log(v)} \frac{1}{n}$$

We used it with *birth rate* $b_n = b \cdot n$, *death rate* $d_n = d \cdot n$ and *immigration rate* $b_0 = m$, with m having measured experimentally, $m = 0.05$.

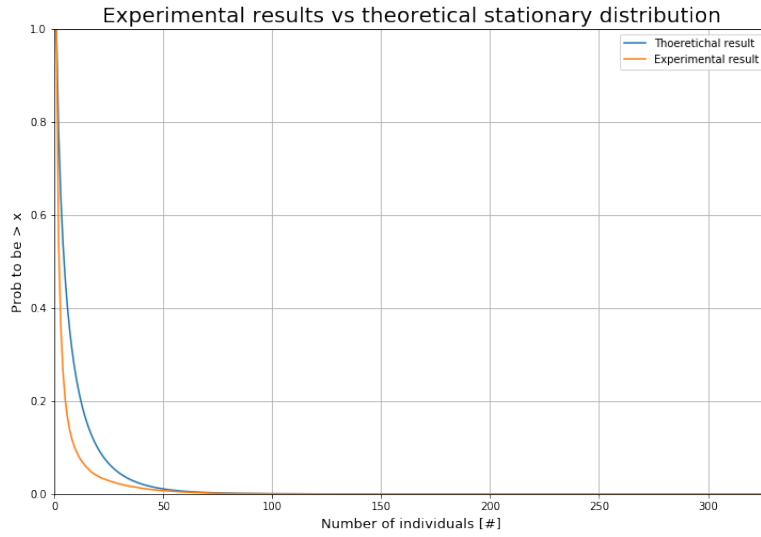


Figure 2: Figure shows the comparison between empirical and theoretical results.

As Figure 2 shows the theoretical results are generally bigger than the experimental ones. This could be due to the way we divided the subplots (because we after divided them we treated them as independent).

Task 4

In this section we evaluated the Species Area Relationship (SAR). For this purpose we choose some area values for the subsets in which the environment is divided; then we evaluated the mean number of species averaging over all the subplots. We also used the formula obtained theoretically for the average number of species with respect to the area of the subplot:

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

and fit it finding the best m parameter that describes the data. The area was divided in square subplots of different sizes; for each plot the number of species was evaluated and then the average was taken over all subplots. We used the following areas: [16, 25, 100, 400, 625, 2500, 10000, 15625, 62500] which corresponds to square subplots whose side is respectively: [4, 5, 10, 20, 25, 50, 100, 125, 250]. Our fit result is: $m_{fit} = 0.0027 \pm 0.0004$, while the expected result is: $m_{given} = 0.05$

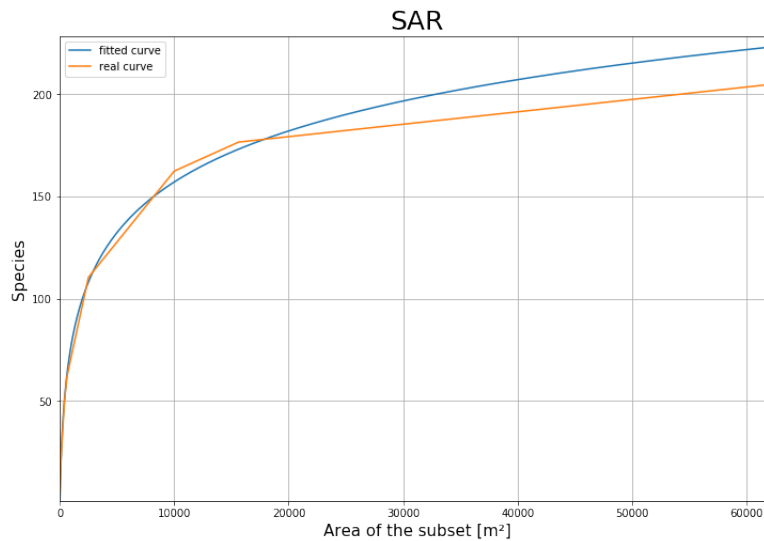


Figure 3: Figure shows the experimental and the theoretical results for the Species Area Relationship.

Task 5

In the last point we build the *logistic growth model* starting from the master equation. In the *logistic growth model* species abundance initially grows exponentially but then reaches a maximum value due to the carrying capacity K of the system.

To gain it we use:

- Birth: $A+0 \rightarrow A+A$ with probability b (per unit of time and per individual);
- Migration: $0 \rightarrow A$ with probability m (per unit of time and per individual);
- Death: $A \rightarrow 0$ with probability $d(A) = b \frac{A}{K}$ (per unit of time and per individual).

We can write:

$$W^+ = W(A \rightarrow A+1) = b \cdot A + m$$

$$W^- = W(A \rightarrow A-1) = \frac{b \cdot A}{K} \cdot A$$

Writing the Master Equation

$$\frac{dP_n}{dt} = P_{n+1}W_{n+1}^- + P_{n-1}W_{n-1}^+ - P_n \cdot (W_n^+ + W_n^-)$$

Now we define the current probability as:

$$J(n) = P_n W_n^- - P_{n-1} W_{n-1}^+$$

So the Master Equation becomes:

$$\frac{dP_n}{dt} = J(n+1) - J(n)$$

Imposing stationarity we find:

$$0 = J(n+1) - J(n) \Rightarrow J(n+1) = J(n)$$

We also have:

$$J(0) = 0 \Rightarrow P_0 W_0^- - P_{-1} W_{-1}^+ = 0$$

Using following boundaries:

$$W_{-1}^+ = 0 \quad (\text{this is not a physical situation})$$

$$W_0^+ = 0 \quad (0 \text{ individuals} \rightarrow 0 \text{ births})$$

we have:

$$J(0) = 0 \Rightarrow J(n+1) = 0 \quad \forall n$$

It holds then:

$$J(n+1) = P_{n+1} W_{n+1}^- - P_n W_n^+ \Rightarrow P_{n+1} = \frac{W_n^+}{W_{n+1}^-} P_n$$

So, $\forall P_n$, it holds:

$$P_n = \left(\prod_i^n \frac{W_i^+}{W_{i+1}^-} \right) P_0$$

Now, to find P_0 we have to normalize, so:

$$\sum_{n=0}^K P_n = 1 \Rightarrow P_0 \sum_{n=0}^K \frac{P_n}{P_0} = 1$$

Using the fact that all fractions written above are not dependent on P_0 , we can write:

$$P_0 = \frac{1}{\sum_{n=0}^K \left(\prod_{i=0}^n \frac{W_i^+}{W_{i+1}^-} \right)}$$

So we evaluated the stationary probabilities P_n^* in this way and plotted them solution in this terms, using the following parameters (as examples):

μ_{exp}	μ_{fit}	$\mu_{fit}error$
8.95	8.74	0.04
σ_{exp}	σ_{fit}	$\sigma_{fit} error$
3.20	3.18	0.04

- $m = 0.1$
- $b = 1$
- $K = 10$

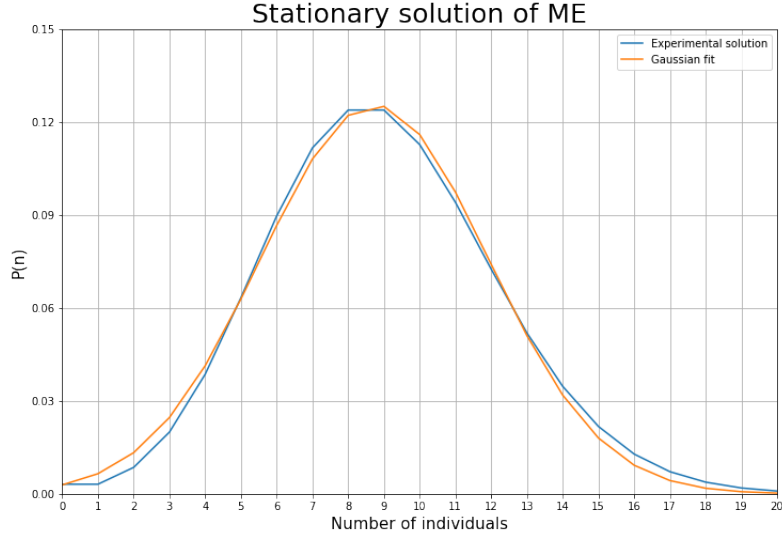


Figure 4: Figure shows the distribution of stationary probabilities for the master equation.

The Figure 4 shows the probability distribution of the stationary solution of the master equation. We evaluated the mean μ and standard deviation σ of the stationary population. We then decided to fit the stationary distribution using a gaussian. The results are the following:

- $\mu = 8.74 \pm 0.04$
- $\sigma = 3.18 \pm 0.04$

We also calculated the Kullback-Leibler divergence between our probability distribution and the gaussian we fit, finding $D(P_{exp}||Gauss) = 0.01265$. This value indicates that the 2 distributions are quite similar; this allows us to have a good estimation of the value that maximizes the distribution, which is the μ of the gaussian itself. This value represents our best estimation of the number of individuals present at the stationary regime.