Neutral theory and emergent patterns in the BCI dataset

Andrea Perin, 1206078, Carlo Polato, 1205688, and Vincenzo Maria Schimmenti, 1204565

(Dated: 14 May 2019)

In this exercise, the BCI dataset is once again analyzed, with a focus on the predictions of Neutral Theory. In particular, the stationary solution of a birth-death Master Equation will be compared with the empirical cumulative *Relative Species Abundance (RSA)*, and then the empirical *Species Area Relationship (SAR)* will be compared to the theoretical predictions of Neutral Theory for a well-mixed population. Finally, a microscopic derivation of logistic growth will be presented, and its dynamics studied.

I. PRELIMINARIES

The BCI dataset was initially pruned of its non relevant entries, namely those corresponding to dead trees. The resulting dataset contains a total of $S = 299$ different species. The 50 hectare plot was then divided in $N = 800$ square subplots of side $l = 25m$ and area $a = 625m^2$. Each of these subplots was supposed to be independent, as to make calculation of statistical quantities feasible.

II. TASKS 1-2: VECTOR OF ABUNDANCES AND EMPIRICAL RSA

For each of the N subplots and for each of the S species, the number of individuals was collected, thus building a $N \times S$ matrix $\mathbf{M}$, the entries of which are such that:

$$(\mathbf{M})_{ij} = \text{number of individuals of species } j \text{ in subplot } i$$

As most species have a low number of individuals, $\mathbf{M}$ is expected to be rather sparse. Instead of using it, then, the vector $\mathbf{X}$ was built, by flattening the matrix after its 0-valued elements had been eliminated. $\mathbf{X}$ contains all the info on present species in the whole plot. The empirical cumulative distribution function was then built from $\mathbf{X}$; that is, the probability that, picking a species at random, it has an abundance that is greater or equal than a given value x. This actually amounts to calculating the cumulative *Relative Species Abundance (RSA)*, $P_{\geq}(x)$, which is depicted in Figure 1.

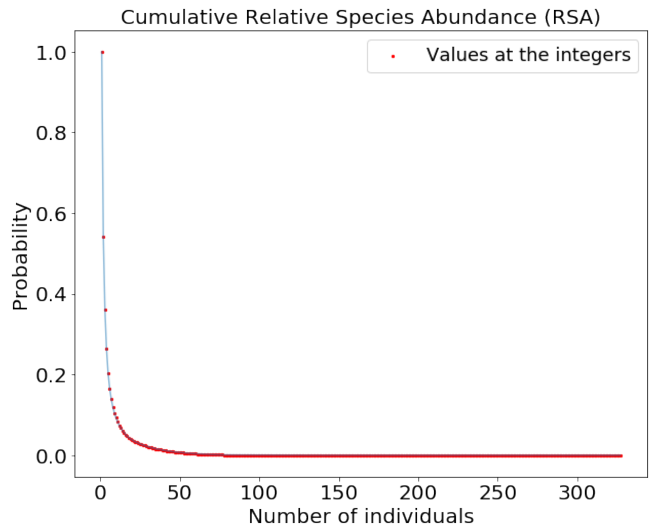

FIG. 1. The plot of the cumulative RSA $P_{\geq}(x)$ as a function of the number of individuals. Of course, $P_{\geq}(1) = 1$.

III. TASK 3: COMPARISON OF $P_{\geq}(x)$ WITH A BIRTH-DEATH MASTER EQUATION STATIONARY SOLUTION

As the title implies, the cumulative RSA previously computed was compared with the solution of a birth-death Master Equation, the rates of which are as follows:

$$\begin{cases} b(n) = b \cdot n \quad \text{for } n \geq 1, \\ b(0) \equiv \nu = 0.05, \\ d(n) = d \cdot n \end{cases}$$

The stationary solution of this birth-death master equation, with N as the maximum number of possible individuals, is

$$P_n^\star = \frac{N}{n} P_0^\star (1-\nu)^{n-1}$$

where the value of $P_0^\star$ can be retrieved by using the normalization

$$\sum_{n=1}^{\infty} P_n^\star = 1$$

which leads to $P_0^\star = \frac{\nu - 1}{N \log(\nu)}$, and so

$$P_n^\star = -\frac{(1-\nu)^n}{n \log \nu}$$

Denoting the cumulative distribution of $P_n^\star$ by $P_{\geq n}^\star$, that is,

$$P_{\geq n}^\star = \sum_{i \geq n}^{\infty} P_i^\star,$$

the comparison between $P_{\geq n}^\star$ and the empirical cumulative RSA, $P_{\geq}(n)$, is shown in Figure 2.

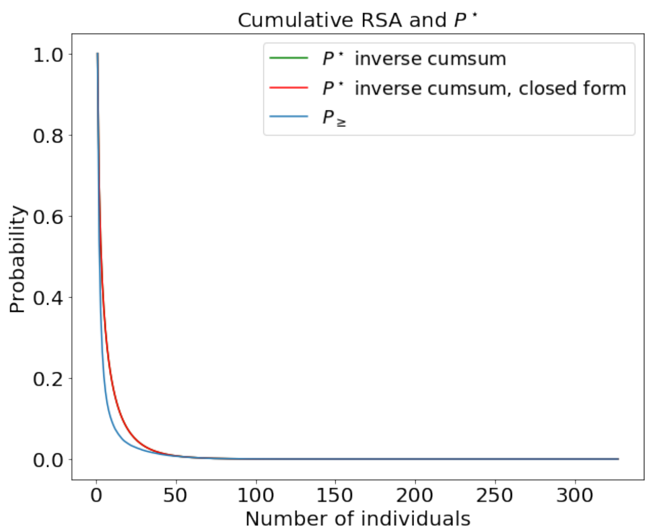

FIG. 2. The plot of $P_{\geq}(x)$ and $P^{\star}_{\geq n}$ as a function of the number of individuals. Both the closed form (corresponding to the sum up to ∞) and the empirical one are plotted, but no visible difference can be perceived: the two overlap almost perfectly.

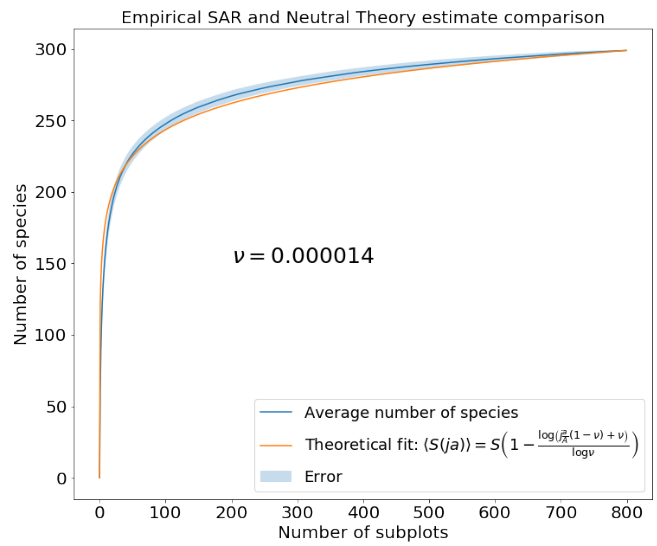

FIG. 3. The empirical SAR, together with its std, and the fitted theoretical counterpart.

IV. TASK 4: EMPIRICAL SPECIES AREA RELATIONSHIP (SAR) AND THEORETICAL PREDICTION

The aim of this task was to compare the empirical *Species Area Relationship (SAR)* and the corresponding theoretical prediction from Neutral Theory, fitting the latter to the former. To do so, the empirical SAR was computed as $\langle S(j\frac{a}{A})\rangle$, where $S(\cdot)$ denotes the number of species present in a given area, a is the area of a single subplot, A is the area of the whole plot, and $j = 1, \dots N$ denotes the number of subplots taken into account (so that $N = 800$). The average $\langle\cdot\rangle$ is understood to be an average over multiple realizations: the value of $S(j\frac{a}{A})$ depends on what the j subplots are, so that the order according to which subplots are picked is important. By averaging over multiple orderings, $\langle S(j\frac{a}{A})\rangle$ can be regarded as a meaningful estimate of a "subplot-independent" SAR.
The theoretical prediction, on the other hand, yields the following result:

$$\left\langle S\left(j\frac{a}{A}\right)\right\rangle = S\left(1 - \frac{\log\left(j\frac{a}{A}(1-\nu)+\nu\right)}{\log\nu}\right)$$

As the ratio $\frac{a}{A} = \frac{1}{N}$ was fixed at the beginning, the only free parameter that can be employed for the fit is ν.
A direct fit yielded $\nu \approx 0.000014$. The resulting function is shown, together with the empirical SAR, in Figure 3.

V. TASK 5: A MICROSCOPIC DERIVATION OF LOGISTIC GROWTH

The last task was to derive a logistic growth model from a microscopic approach.Given the carrying capacity, K, and the number of individuals, A, let us define the following events:

- birth: $A \to A + A$, w.p. b (per unit of time and per individual)
- migration: $0 \to A$, w.p. m (per unit of time)
- death: $A \to 0$, w.p. $d(A) = b\frac{A}{K}$ (per unit of time and per individual)

The transition rates are:

$$W(A \to A+1) \equiv W^{+}(A) = bA + m$$
$$W(A \to A-1) \equiv W^{-}(A) = b\frac{A^2}{K}$$

Bearing these in mind, the Master Equation can be written as:

$$\begin{aligned}\frac{dP_n(t)}{dt} =& P_{n-1}(t)W^{+}(n-1) + P_{n+1}(t)W^{-}(n-1) \\ &- P_n(t)\left(W^{+}(n) + W^{-}(n)\right)\end{aligned}$$

For the stationary solution, $P^{\star}_n$, the Master Equation reads:

$$\begin{aligned}0 =& P^{\star}_{n-1}W^{+}(n-1) + P^{\star}_{n+1}W^{-}(n-1) \\ &- P^{\star}_n\left(W^{+}(n) + W^{-}(n)\right)\end{aligned}$$

Denoting

$$J_n \equiv P^{\star}_n W^{-}(n) - P^{\star}_{n-1}W^{+}(n-1),$$

then the Master Equation for the stationary solution can be written as

$$J_{n+1} - J_n = 0 \quad \forall n$$

which implies

$$\sum_{x=0}^{n} (J_{x+1} - J_x) = 0 \qquad \Longrightarrow \; J_{n+1} - J_0 = 0$$

Since birth is not properly defined for a negative number of individuals, we can take $W^+_{-1} \equiv 0$, so that

$$J_0 = P^\star_0 W^-_0 - P^\star_{-1} W^+_{-1} = 0$$

and so

$$J_{n+1} = 0 \implies P^\star_{n+1} W^-_{n+1} - P^\star_n W^+_n = 0$$

leading to the recursive expression

$$\begin{cases} P^\star_{n+1} = \frac{W^+_n}{W^-_{n+1}} P^\star_n, & n \geq 0 \\ P^\star_0 = \alpha \end{cases}$$

where α can be calculated by imposing the normalization condition:

$$\sum_{n=1}^{\infty} P^\star_n = 1$$

We get:

$$P^\star_n = \frac{W^+_{n-1}}{W^-_n} P^\star_{n-1} = \frac{W^+_{n-1} W^+_{n-2} \cdots W^+_0}{W^-_n W^-_{n-1} \cdots W^-_1} P^\star_0,$$

Setting our rates:

$$P^\star_n = \prod_{i=1}^{n} \left(1 + \frac{r-1}{i}\right) \frac{K^n}{n!} P^\star_0,$$

where $r \equiv \frac{m}{b}$.
The normalization condition is:

$$(P^\star_0)^{-1} = 1 + \sum_{n=1}^{\infty} P^\star_n$$

The values of the stationary solution for the various values of n are shown in Figure 4.
For reasons unbeknownst to us, the results of the Gillespie simulations of the system do not agree with the theoretical stationary distribution. It appears that, while the shapes are in good agreement, they are not centered around the same mean value (which is slightly above $K = 10$ for the simulations and slightly below for the distributions).

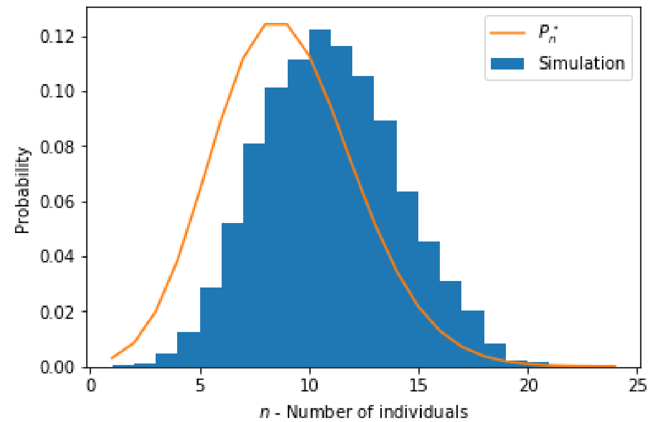

FIG. 4. The stationary distribution of the solution for the Master Equation for $K = 10$.

A linear noise approximation

A simplified version of Van Kampen's *linear noise approximation (LNA)* can be employed in order to retrieve a continuous stochastic approximation (cfr. 'A simplified derivation of the Linear Noise Approximation' here). By employing this method, the following equations were obtained:

$$\begin{cases} \dot{x} = bx - bx^2 + \frac{m}{K} \\ \dot{\xi} = b\,(1 - 2x_{eq})\,\xi + \sqrt{b^2 + 4b^2 x_{eq}^2}\,\eta \end{cases}$$

where x denotes the fraction of individuals over the carrying capacity, ξ is a noise component, described by the SDE above (where η is a white noise) and x_{eq} is the equilibrium position of x; approximately, the number of individuals A is related to x and ξ via:

$$A(t) \approx Kx(t) + \sqrt{K}\xi(t)$$

The interesting equation is actually the first one, as it reproduces the logistic dynamics one expects from this microscopic derivation. The second, instead, describes the effects of stochasticity; the sum of the two allows the modeling of a stochastic logistic growth, which is depicted in Figure 6.

A. An alternate derivation of logistic growth

One can retrieve the equations of logistic growth by slightly changing the way death rate is interpreted[1]. If death was to be thought of as the result of competition between individuals, the following holds:

$$W(A \to A+1) \equiv W^+(A) = bA$$
$$W(A \to A-1) \equiv W^-(A) = b\frac{A(A-1)}{K}$$

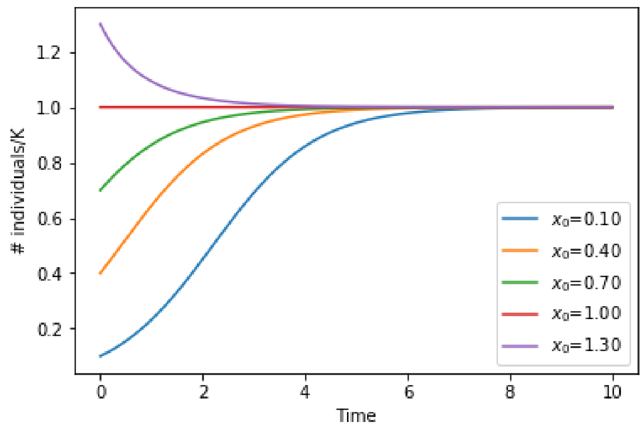

FIG. 5. The deterministic part of the solution of Master Equation for $K = 10$ (LNA).

that is, as long as at least 2 individuals are present, there is the possibility of one of them dying as a result of competition. If, instead, only one individual is present, no death can occur. By using these rates, together with the recursion formulae, one gets:

$$P_n^{\star} = \frac{1}{e^K - 1} \frac{K^n}{n!}$$

which closely resembles a Poisson distribution. No dependence on the parameter b is to be found.

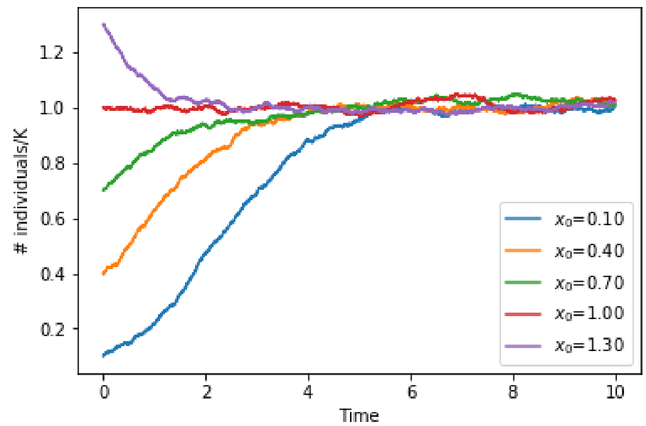

FIG. 6. The deterministic and stochastic part of the solution of the Master Equation for $K = 10$ (LNA).

[1] Cfr. this article.