EXERCISES (i) - Maths for Biology

Computational Methods in Ecology and Evolution
Imperial College London
Silwood Park

Course 2017-18

Lecturer: Alberto Pascual-García.

Correspondence: alberto.pascual.garcia@gmail.com

```
Abbreviations:
 [EXM] This is a type of exercise that you may find in the exam.
 [DLV] This is an exercise that you must hand in.
```

Exercises Modelling

Exercise 1 - Contact matrices and complex networks [DLV] This is an exercise that I expect you to develop along the two weeks, because you will need to know a little bit more about the Snake Puzzle to solve it. For the moment, download from Blackboard in Readings/Teaching the article that refers to the Snake Puzzle (Nido *et al.*) Read the article at least until the section "The contact matrix" (included). The solutions the paper is referring to, can be download in the folder OtherMaterial as "Contact_Matrices.zip". As you can see in Fig. 3, some of the solutions are redundant, so you will just need to work with *half of the solutions*.

In this exercise, the first thing you should do is to transform the solutions from matrix format to a file with the following format:

Monomer1 Monomer2 Contact

Monomer1 Monomer3 Contact

...

MonomerN-1 MonomerN Contact

Where contact is the value that both monomers have in the matrix (either zero or one). Our aim is to transform these matrices into a network, and then to analyse its properties. Therefore, once you have the matrix in this format, you can skip all the rows that have a value of contact equal to zero, as they will not generate any link. The next step is to download and install Cytoscape in your computer (cytoscape.org). I will show you how to import the matrices and work with them.

You will deliver:

- Four networks, one for each solution and cool layout (explore the tab layout). Note that the links are undirected and that it should not have self-loops.

- You will explore network properties using Tools > Network analyzer > Network analysis > analyse network. You will get a lot of information from this analysis. You should try to find any interesting property (one for which you find differences between the solutions), export the graphic(s) and relate them with the properties explained in the manuscript. You may want to see what happens if you delete the links with the nearest neighbours.

Exercises Functions

General questions

Exercise 1 - Function domain. [EXM] Find the domain of the following functions:

1.
$$h(x) = \sin\left(\frac{x+1}{x-1}\right)$$

2.
$$m(x) = \log\left(\frac{x^2+x+3}{x^2+1} - 1\right)$$

Exercise 2 - Function parity. [EXM] Determine the parity of the following functions:

1.
$$f(x) = \frac{\sin(x) + x^3}{2x^2 + \cos(x) + 4}$$

2.
$$g(x) = \frac{3x^4 + x^2}{x^5 + 1}$$

3.
$$n(x) = \frac{\sin(x)\cos(x)}{x}$$

Exercise 3 - Function periodicity. [EXM] Determine if the following functions are periodic:

1.
$$f(x) = \sin(x) + \cos(x)$$

2.
$$m(x) = \frac{\sin(x)}{\cos(x)}$$

Polynomial functions

Exercise 1 - Guessing the extremes of polynomial functions [EXM]. Consider polynomials of your choice fulfilling the following generic form

$$y = \sum_{n=0}^{N} a_n x^n$$

and explore, numerically, if the following sentences "seem to be"[1] true:

- If $a_N > 0$ (the leading coefficient is positive), $y \to \pm\infty$ when $x \to \pm\infty$ if N is even.
- If $a_N > 0$, $y \to \mp\infty$ when $x \to \pm\infty$ if N is odd.
- If $a_N < 0$ whatever behaviour you found to be true now seems to be the opposite.
- All even-degree polynomials behave, on their ends, like quadratics, and all odd-degree polynomials behave, on their ends, like cubics.

According with the above results, fill in Table 1 the expected behaviour for the variable y for a generic polynomial.

[1] We say "seem to be" because you are not able to consider infinite vectors in your numeric tests.

Polynomials with	Positive leading coefficient	Negative leading coefficient
Odd degree	$x \to \infty$ $y \to ?$ $x \to -\infty$ $y \to ?$	$x \to \infty$ $y \to ?$ $x \to -\infty$ $y \to ?$
Even degree	$x \to \infty$ $y \to ?$ $x \to -\infty$ $y \to ?$	$x \to \infty$ $y \to ?$ $x \to -\infty$ $y \to ?$

Table 1: Exercise 1 of polynomial functions.

Exercise 2 - Finding the roots of polynomial functions [EXM] (Fundamental Theorem of Algebra). We start now stating the FUNDAMENTAL THEOREM OF ALGEBRA "A polynomial of order N has, counted with multiplicity, exactly N complex roots which are the solutions of $f(x) = 0$".

In particular, we know that, for a second degree polynomial $f(x) = ax^2+bx+c$, its roots can be found applying the formula:

$$x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}.$$

Taking into account the results found in Exercise 1 and this theorem, plot (approximately) the following two functions without the help of your computer: $y_1 = x^2 - 3x + 2$ and $y_2 = x^3 + 2x^2 - 3x$. Then, test with your computer the accuracy of your guess.

Exercise 3 - The multiplicity of the roots [EXM] (Fundamental Theorem of Algebra). In mathematics, we will call *multiplicity* to the number of times that a number appears in a set of numbers. For instance, in the set of numbers $\{4, 5, 6, 6, 7\}$ the number six has multiplicity 2 and all the other numbers have multiplicity one. Therefore, if we consider the set of roots of a polynomial, it might happen that one of these roots appears more than once. As polynomials are functions and we know that there is only one value of y for a given value of x, if x is a root of the polynomial and it has, for instance, multiplicity equal to two, we can imagine that the function is taking the value $y = 0$ "twice" at that point[2], and we count it as a "double" root.

For instance, consider this function $y_3 = x^4 + 2x^3 - 3x^2$. According with the above theorem we know that it has four roots. However, there is a root $x = 0$, which we say it has multiplicity 2 because, if we perform the factorization $y_3 = x^2(x^2 + 2x - 3)$, which leads to the two equations:

$$x^2 = 0$$
$$x^2 + 2x - 3 = 0,$$

we observe that the first equation would lead to the solution $x = \pm\sqrt{0}$, i.e. $x = 0$ twice. In order to plot roots with multiplicity larger than one, we should take into account these two rules:

- Any roots with an even multiplicity (twice, four times, six times, etc) are squares, so the function doesn't change the sign. This means that the function "touches" zero and "come back" to the same quadrant.
- Any root with an odd multiplicity will cross the x-axes, because it will change the sign.

Taking into account these considerations, plot "by hand" the function y_3 and the functions $y_4 = (x + 3)^2(x - 2)$ and $y_5 = (x + 3)^2(x - 2)^2$. Then, test with your computer the accuracy of your guess.

Fractional functions

Exercise 1. Fractions decomposition [EXM]. Decompose in simple fractions the following rational functions:

1.

$$f(x) = \frac{3x}{x^2 - 6x + 8}$$

[2] Of course this picture is not rigurous but we are not mathematicians so we don't care much. We could also say that this root is "degenerate". This term is widely used in Physics to say that more than one state of our system is described by the same values of the variables we are considering to describe it. For instance, two configurations of the system may have the same energy, and we will say that this value of the energy is degenerated.

Trigonometric functions

Exercise 1. Common trigonometric functions Plot with your computer the functions:

$$y_1 = \sin(x)$$
$$y_2 = \sin(x + \pi/2)$$
$$y_3 = \cos(x)$$
$$y_4 = \cos(x + \pi/2)$$
$$y_5 = \frac{\sin(x)}{\cos(x)} = \tan(x)$$
$$y_6 = \frac{\cos(x)}{\sin(x)} = \cot(x)$$

Determine the domain, image and period of the functions. Which of them is bounded?

Exponential functions

Exercise 1. Growing and dumping Exponential functions have the form $y = A\exp(kx)$, where A is the amplitude and k, if positive, can be interpreted as a growth rate while, if negative, as a dumping term. Plot with your computer in the same graphic the following functions:

$$y_1 = \exp(x)$$
$$y_2 = \exp(-x)$$
$$y_3 = 10^x$$
$$y_4 = 10^{-x}$$

In these examples we fixed $A = 1$ and $|k| = 1$, and we changed the base. Repeat now the exercise for different values of A and k. Determine the domain and image. Are they bounded?

Exercise 2. Radioactive decay [DLV]. The amount of radioactive carbon-14 that remains in a fossil a time t after death is given by the function:

$$C = C_0 \exp(-kt)$$

where k is a positive constant and C_0 is the amount of carbon-14 contained in the body before death. Living beings aquire carbon during their life and this quantity should be estimated for each organism, assuming that the concentration of carbon in the biosphere is roughly constant (an assumption that could be wrong for different reasons such as the existence of volcanic activity). We will keep it as an unknown constant and, knowing that the half-time of carbon-14 (the time elapsed to reduce the initial amount of carbon to half its value) is ~5370 years, you should be able to find the value of the constant k (and its dimensions).

Hyperbolic functions

Exercise 1. Common hyperbolic functions Plot the following functions with your computer:

$$y_1 = \sinh(x)$$
$$y_2 = \cosh(x)$$
$$y_3 = \frac{\sinh(x)}{\cosh(x)} = \tanh(x)$$
$$y_4 = \frac{\cosh(x)}{\sinh(x)} = \coth(x)$$

Exercise 2. Approximating hyperbolic functions [DLV] Consider the following functions:

$$y_5 = x + \frac{1}{2!}x^2$$
$$y_6 = x + \frac{1}{3!}x^3$$
$$y_7 = \exp(x)/2$$
$$y_8 = \exp(-x)/2$$
$$y_9 = -\exp(-x)/2$$

which can be considered a good approximation of the functions $\cosh(x)$ and $\sinh(x)$, respectively, for some values of x. To explore in which region the approximation is justified, we could use the Root Mean Square Deviation:

$$RMSD(n) = \frac{1}{n}\sqrt{\sum_{x=0}^{n} (f(x) - g(x))^2}$$

where f and g are functions, and we should note that the $RMSD$ is a function of n. Plot in a single graphic the function $RMSD(n)$ versus n for the following cases:

- For $f(x) = y_1$ and $g(x) = y_5$ when $n \in [0, 10]$.
- For $f(x) = y_2$ and $g(x) = y_6$for $n \in [0, 10]$.
- For $f(x) = y_1$ and $g(x) = y_7$ when $n \in [0, 10]$.
- For $f(x) = y_2$ and $g(x) = y_7$for $n \in [0, 10]$.
- For $f(x) = y_1$ and $g(x) = y_8$ when $n \in [-10, 0]$.
- For $f(x) = y_2$ and $g(x) = y_9$for $n \in [-10, 0]$.

and determine which is the region where the approximation seems to be reasonable for each case.

Logarithmic functions

Exercise 1. Logarithmic functions Plot the following functions with your computer:

$$y_1 = \ln(x)$$
$$y_2 = -\ln(x)$$
$$y_3 = \log(x)$$
$$y_4 = \log_2(x)$$

Determine the domain and image of the functions.

Exercise 2. The True Diversity [DLV]. An important information-theoretic measure of a discrete probability distribution (of course there is a continuous version as well) of N independent events is the Shannon entropy:

$$S = -\sum_{i}^{N} p_i \log_2(p_i)$$

where p_i is the probability of event i. It can be interpreted as the amount of information (in bits) we gain when we run one experiment and we obtain that the event i has happened. Equivalently, it can be seen as the amount of uncertainty (in bits) we had before the experiment was run. You should remind this apparently innocent difference in the way we interpret the informational entropy, because you will find a similar epistemological challenge when we interpret probabilities in a *frequentist* or *bayesian* framework[3], as you will see in other courses. This measure has been widely used in ecology to analyse the distribution of species. Instead, the natural logarithm is used (Shannon index, H) and the effective number of different types of species, which has been called the true diversity D, is simply computed as $D = e^H$.

Now consider three different populations with $N_1 = 10$, $N_2 = 50$ and $N_3 = 100$ species, respectively and then:

- Compute, numerically, the Shannon entropy S, the Shannon index H and the True diversity D for the three populations assuming that the total number individuals of *all* species in each population is distributed among the different species $i = 1, .., N$, as:
  - i) uniformly, i.e. $P(x = i) = 1/N$

[3] http://stats.stackexchange.com/questions/22/bayesian-and-frequentist-reasoning-in-plain-english#56

– ii) as $P(x = i) = \frac{1}{Z}\frac{1}{2^i}$ with $Z = \sum_i^N \frac{1}{2^i}$.

- Present your results in a table, and explain them considering the two interpretations of the Shannon entropy highlighted above.
- Check the following Ref. [1] and, taking into account what you learned here, chat with your mates in the bar about entropy and consciousness (show evidence with a picture).

Some final important operations

Exercise 1 - Function superposition. Plot in the same graph the following functions:

$$f_1(x) = -2x$$
$$f_2(x) = x^3$$
$$y = f_1(x) + f_2(x)$$

Exercise 2 - Function modulation Plot in the same graph the following functions:

$$f_1(x) = \sin(x)$$
$$f_2(x) = -\sin(x)$$
$$y = f_1(x)f_2(x)$$

Have a look to these articles:

1. https://en.wikipedia.org/wiki/Signal-to-noise_ratio.
2. Frequency Modulation (Wikipedia).And look for a relationship with the previous article.
3. Ref. [2].
4. Chat in the bar with your mates about cancer, frequencies and modulation.

References

[1] R. G. Erra, D. Mateos, R. Wennberg, and J. P. Velazquez, "Statistical mechanics of consciousness: Maximization of information content of network is associated with conscious awareness," *Physical Review E*, vol. 94, no. 5, p. 052402, 2016.

[2] J. W. Zimmerman, M. J. Pennison, I. Brezovich, N. Yi, C. T. Yang, R. Ramaker, D. Absher, R. M. Myers, N. Kuster, F. P. Costa, *et al.*, "Cancer cell proliferation is inhibited by specific modulation frequencies," *British journal of cancer*, vol. 106, no. 2, pp. 307–313, 2012.