Evolutionary Dynamics

Assignment #01

Aliaksandr Yudzin, Lorenzo Gatti

Wednesday 3rd October, 2012

1.1 Problem 1: Logistic difference equation

We can find stable points from

$$f(x, a) = ax(1 - x) = x \tag{1.1}$$

as the roots of the quadratic equation:

$$ax^2 - (a - 1)x = x(ax - (a - 1)) = 0 \tag{1.2}$$

which are:

$$x_1 = 0 \tag{1.3}$$

$$x_2 = \frac{(a - 1)}{a} \tag{1.4}$$

The fixed point x_2 is non-negative if $a \geq 1$. One can analyze the local stability of the difference equation 1.1 by examining the partial derivative of f with respect to x evaluated at each fixed point x^*:

$$f' = -ax - a(x - 1) \tag{1.5}$$

Substituting x_1 and x_2 into 1.5 yields:

$$f'(x_1) = a \tag{1.6}$$

$$f'(x_2) = 2 - a \tag{1.7}$$

One finds that if $a > 1 \rightarrow f'(x_1) > 1$ with $x_1 = 0$, then x^* is repelling, while if $1 < a < 3 \rightarrow 1 > f'(x_2) > -1$ with $x_2 = (a - 1)/a$ is stable (attractor).

1.1.1 Point stability at different values of a

When $a = 0.9$ then $x_1 = 0.9, x_2 = 1.1$
When $a = 2.1$ then $x_1 = 2.1, x_2 = -0.1$

For a values in excess of 3.57, the orbits $x(t, x0) = x0, x1, x2, \ldots$ depend crucially on the initial condition $x0$. Slight variations in $x0$ result in dramatically different orbits, an important characteristic of chaos.
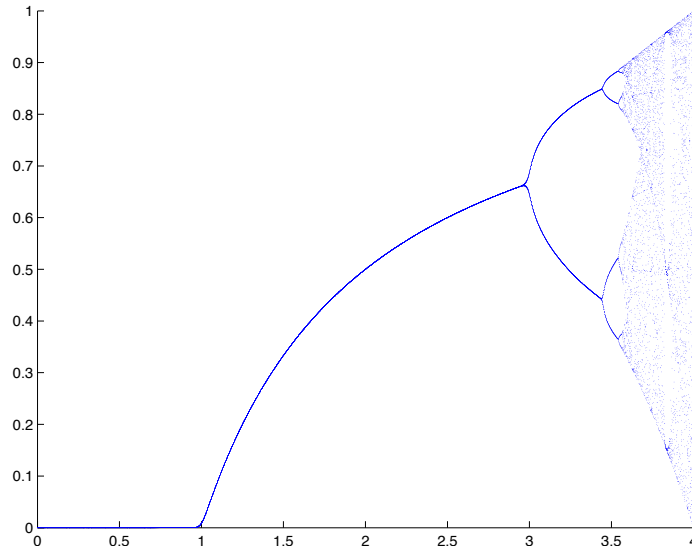
Figure 1.1: Logistic Map Bifurcation Diagram

2.1 Problem 2: Logistic growth in continuous time

We have to solve the equation

$$\frac{dx}{dt} = rx(1 - \frac{x}{K}) = r(x - \frac{x^2}{K}) = \frac{rx(K - x)}{K} \tag{2.8}$$

using the separation of variables

$$\frac{K}{x(K - x)}dx = rdt \tag{2.9}$$

decomposing the right part with partial fractions

$$\frac{K}{x(K - x)} = \frac{A}{x} + \frac{B}{K - x} \tag{2.10}$$

We find A and B

$$A = \frac{K - Bx}{K - x} \tag{2.11}$$

$$B = \frac{K - A(K - x)}{x} \tag{2.12}$$

Supposing $A = 1$, and according to above, B is also equal to 1, so our partial fraction decomposition is

$$\frac{K}{x(K - x)} = \frac{1}{x} + \frac{1}{K - x} \tag{2.13}$$

Now we have to take the integral from both parts:

$$\int \frac{1}{x} + \frac{1}{K-x} dx = \int r dt \tag{2.14}$$

$$\int \frac{1}{x} + \int \frac{1}{K-x} dx = r \int dt \tag{2.15}$$

$$\ln x - \ln K - x = rt + x_0 \tag{2.16}$$

$$\ln \frac{x}{K-x} = rt + x_0 \tag{2.17}$$

$$\frac{x}{K-x} = x_0 e^{rt} \tag{2.18}$$

Finally the solution

$$x(t) = x_0 K e^{rt} \frac{1}{K + x_0(e^{rt} - 1)} \tag{2.19}$$

2.1.1 Determining the stability of equilibria

To determine the stability of the two equilibria points found solving the equation 2.8 in 0

$$x_1 = 0 \tag{2.20}$$
$$x_2 = K \tag{2.21}$$

one has to derive it for x:

$$f' = -r(\frac{x}{K} - 1) - \frac{rx}{K} \tag{2.22}$$

Substituting x_1 and x_2 into 2.22 yields:

$$f'(x_1) = -1 \tag{2.23}$$
$$f'(x_2) = -r \tag{2.24}$$

One finds that if $f'(x_1)$ with $x_1 = 0$, then x^* is always stable, while if $for r < 0 \rightarrow f'(x_2) > 0$ with $x_2 = -r \; x^*$ is repelling.

3.1 Problem 3: Hardy-Weinberg equilibrium

Assuming HWE fot three alleles (allele and genotype frequences):

$$(A + B + 0)^2 = A^2 + B^2 + 0^2 + 2AB + 2A0 + 2B0 = 1$$

We could solve the equation:

$$(A + B + 0)^2 = (p + q + r)^2 = p^2 + q^2 + r^2 + 2pq + 2pr + 2qr = 1$$

And given observable genotype frequences

$$0^2 = \frac{900}{10000} = 0.09 = r^2 \rightarrow r = 0.3 \tag{3.25}$$

$$2AB = \frac{2000}{10000} = 0.20 = 2pq \rightarrow pq = 0.1 \tag{3.26}$$

$$A^2 + 2A0 = \frac{1600}{10000} = 0.16 = p^2 + 2pr \tag{3.27}$$

$$\tag{3.28}$$

$$p^2 + 2*0.3*p = 0.16 \tag{3.29}$$

$$p_1 = \frac{1}{5} \tag{3.30}$$

$$p_2 = -\frac{4}{5} \text{ not acceptable} \tag{3.31}$$

$$\tag{3.32}$$

$$q = \frac{0.1}{0.2} = 0.5 \tag{3.33}$$

So, the frequency for B allele is equal to 0.5

3.1.1 HW Equilibrium

$$A^2 = \frac{1500}{10000} = 0.15 = p^2 \rightarrow p = 0.39 \tag{3.34}$$

genotype	expected	observed
PP	10000*0.04 = 400	1500

Hence, assuming population carrying AA is 1500, we are not at the equilibrium of HW ($p_b^2 = 0.15 \neq p_a^2 = 0.04$)

4.1 Problem 4: Sequence alphabets

4.1.1 A: Amino acid sequences

According to the number of amino acids c, $c_1 = 20$ standard aminoacids$|c_2 = 22$ standard + non-canonical we consider there are:

$$\text{for } c_1 = 20 \text{ then } s_{space} = c_1^L, 20^{50} = 1.13 * 10^{65} \tag{4.35}$$

$$\text{for } c_2 = 22 \text{ then } s_{space} = c_2^L, 22^{50} = 1.32 * 10^{67} \tag{4.36}$$

Since each aminoacid is not codified by an unique codon (3 nt), the number of unique DNA sequences is much larger then unique aminoacids sequences.

4.1.2 B: Amino acids sequences and DNA sequences

Each amino acid is codified by 3 nt, so for a sequence of 50 amino acids we need 150 nt or 153 nt if we consider the codon stop at the end of the codifying sequence.

$$\text{for } L = 150 \text{ then } d_{space} = c^L = 4^{150} = 2.04 * 10^{90} \tag{4.37}$$

$$\text{for } L = 153 \text{ then } d_{space} = c^L = 4^{153} = 1.30 * 10^{92} \tag{4.38}$$

5.1 Problem 5: Random sequences

5.1.1 a: Average and Expected distance

If the Hamming distance is the number of coordinates where two vectors $x = (x_1, \ldots, x_n)$ and $y = (y_1, \ldots, y_n)$ of length N differ

$$d_H(x, y) = \sum_{i=1}^{n} |x_i - y_i|$$

For a set $A \subseteq F_2^n$, $|A|$ denotes the cardinality of A. The average distance in A is defined by

$$dist(A) = \frac{1}{|A|^2} \sum_{x \in A} \sum_{y \in A} d_H(x, y) \tag{5.39}$$

Hence

$$dist(F_2^n) = \frac{1}{2^{2n}} \sum_{x \in F_2^n} \sum_{y \in F_2^n} d_H(x, y) = \frac{n 2^{n-1} 2^n}{2^{2n}} = \frac{n}{2} \tag{5.40}$$

Let V be some finite set with q elements where $n \geq 1$ and V^n is the dimension of sequence space. P is the common probability distribution. Then, we have

$$\frac{n(q-1)}{q} - L(P) \leq Ed_H(X, Y) \leq \frac{n(q-1)}{q}$$

where $L(P)$ measures how skewly P is distributed as

$$L(P) = q^{n-1} \sum_{x \in V^n} \left[P(x) - \frac{1}{q^n} \right]^2$$

If $2 \leq q \geq 4$, (in DNA $q = 4$) then

$$Dd_H(X, Y) \leq \frac{n(q-1)}{q^2} + \frac{2}{q} L(P)$$

Hence

$$\frac{n(q-1)}{q^2} \leq Ed_H(X, Y) \leq \frac{n(q-1)}{q^2} + \frac{2}{q} L(P)$$

This gives out that for two i.i.d. random sequences with the common probability distribution P, we have for $n = L$ and $q = 4$

$$Ed_H(X, Y) = \frac{n(q-1)}{q} = \frac{L(4-1)}{4} = \frac{3}{4} L \tag{5.41}$$

5

5.1.2 b: Number of sequences at a certain distance

For a binary sequence, it is a composition of n elements in L places at distance 1

$$\frac{L!}{(L-1)!}$$

It is a composition of n elements in L places for K differences:

$$\frac{L!}{(L-K)!}$$

For DNA sequences:

could be a permutation with repetitions?

5.1.3 c: Replication of sequences

$$q_{ij} = p^{H_{ij}}(1-p)^{L-H_{ij}} \tag{5.42}$$

6.1 Problem 6: Quasispecies

$$\frac{dx}{dt} = wx - \Phi x$$

$$w = (w_{ij}) = (f_j q_{ji}) = \begin{bmatrix} 0 & q_{10} \\ f_0 q_{01} & 0 \end{bmatrix}$$

In the limiting case of completely error-free replication, Q becomes the identity matrix: all diagonal entries are one, all off-diagonal entries are zero. In our case the replication is error free with probability q.

$$\text{with } f_1 = 1 \text{ and } f_0 > 1 \rightarrow w = \begin{bmatrix} 0 & q \\ fq & 0 \end{bmatrix}$$

Φ is the largest eigen value of w, while x is the eigen vector of w.

To find eigen values and eigen vector: $|w - \lambda I| = 0$

$$\begin{bmatrix} -\lambda & q \\ fq & -\lambda \end{bmatrix} = 0 \rightarrow \lambda^2 - fq^2 = 0 \rightarrow \begin{cases} \lambda_1 = +q\sqrt{f} \\ \lambda_2 = -q\sqrt{f} \end{cases}$$

Φ is the largest eigen value of $w \rightarrow \Phi = \lambda_1 = q\sqrt{f}$.

$$\begin{pmatrix} x_0' \\ x_1' \end{pmatrix} = \begin{pmatrix} 0 & q \\ fq & 0 \end{pmatrix} \begin{pmatrix} x_0 \\ x_1 \end{pmatrix} - q\sqrt{f} \begin{pmatrix} x_0 \\ x_1 \end{pmatrix} \rightarrow \begin{pmatrix} x_0' \\ x_1' \end{pmatrix} = \begin{pmatrix} qx_1 \\ fqx_0 \end{pmatrix} + \begin{pmatrix} -q\sqrt{f}x_0 \\ -q\sqrt{f}x_1 \end{pmatrix} = \begin{pmatrix} qx_1 - q\sqrt{f}x_0 \\ fqx_0 - q\sqrt{f}x_1 \end{pmatrix}$$

6

$$x_0' = qx_1 - q\sqrt{f}x_0 = f(x_0, x_1) \tag{6.43}$$
$$x_1' = fqx_0 - q\sqrt{f}x_1 = g(x_0, x_1) \tag{6.44}$$

To find the equilibrium points $f(x_0, x_1) = g(x_0, x_1) = 0$

$$\begin{cases} x_0^* = \dfrac{1}{\sqrt{f}}x_1^* \\[2mm] x_0^* = \dfrac{\sqrt{f}}{f}x_1^* = \dfrac{1}{\sqrt{f}}x_1^* \end{cases} \rightarrow x_0^* = \dfrac{1}{\sqrt{f}}x_1^*$$

To study the stability of equilibrium point, we can use jacobian matrix:

$$J = \begin{bmatrix} \frac{\partial f}{\partial x_0} & \frac{\partial f}{\partial x_1} \\[2mm] \frac{\partial g}{\partial x_0} & \frac{\partial g}{\partial x_1} \end{bmatrix} = \begin{bmatrix} -q\sqrt{f} & q \\[2mm] fq & -q\sqrt{f} \end{bmatrix}$$

For stability we need negative eigen value from jacobian matrix:

$$|J - \lambda I] = 0 \rightarrow \begin{cases} tr(J) < 0 \\ det(J) > 0 \end{cases}$$

$$tr(J) = -q\sqrt{f} - q\sqrt{f} = -2q\sqrt{f} < 0$$
$$det(J) = (-q\sqrt{f}) * (-q\sqrt{f}) - fq^2 = fq^2 - fq^2 = 0$$

From the determinant of J matrix we can not conclude any result about stability of steady states.

6.1.1   c

If $f_0 = f_1 = 1 \rightarrow x_0^* = x_1^* \rightarrow$ population genotipe 0 and 1 is the same in whole population at steady state.

6.1.2   d

$$\textbf{If } f_0 \gg 1 \rightarrow x_1^* = \sqrt{f}x_0$$

$$\begin{cases} x_0 \text{ goes extinct} \\ x_1 = \infty \rightarrow \text{ whole population is type 1} \end{cases}$$