ETH

Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Department of Computer Science
ETH Zürich

Evolutionary Dynamics

Assignment #01

Lorenzo Gatti, Zahra Karimadini, Aliaksandr Yudzin

Thursday 4th October, 2012

1.1 Problem 1: Logistic difference equation

1.1.1 a: Finding the equilibrium points

We can find equilibrium 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^*:

$$f' = -ax - a(x - 1) = a - 2ax \tag{1.5}$$

For $|f'_x| < 1$ the equilibrium point is attractive.

$$x_1^* = 0 \rightarrow f'_{(x)} = a - 2a(0) = a \rightarrow |a| < 1 \text{ then } x_1^* \text{ is attractive}$$

$$x_2^* = \frac{(a - 1)}{a} \rightarrow f'_{(x)} = a - 2a + 2 = -a + 2 \rightarrow |2 - a| < 1 \text{ for } 1 < a < 3 \text{ then } x_1^* \text{ is attractive}$$

1.1.2 b: Point stability at different values of a

When $a = 0.9$

$$x_1^* = 0 \rightarrow |a| = |0.9| < 1 \text{ \textbf{attractive}} \tag{1.6}$$

$$x_2^* = \frac{(a - 1)}{a} \rightarrow |2 - 0.9| = |1.1| > 1 \text{ \textbf{not attractive}} \tag{1.7}$$

$$\tag{1.8}$$

When $a = 2.1$

$$x_1^* = 0 \rightarrow |a| = |2.1| > 1 \textbf{ not attractive} \tag{1.9}$$

$$x_2^* = \frac{(a-1)}{a} \rightarrow |2 - 2.1| = |0.1| < 1 \textbf{ attractive} \tag{1.10}$$

$$\tag{1.11}$$

When $a = 3..58$

$$x_1^* = 0 \rightarrow |a| = |3.58| > 1 \textbf{ not attractive} \tag{1.12}$$

$$x_2^* = \frac{(a-1)}{a} \rightarrow |2 - 3.58| = |1.58| > 1 \textbf{ not attractive} \tag{1.13}$$

$$\tag{1.14}$$

In a conceptual view we can see 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.
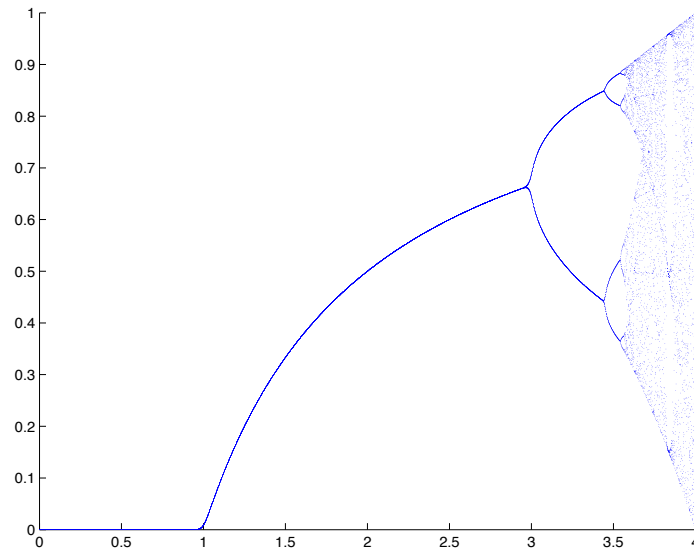

Figure 1.1: Logistic Map Bifurcation Diagram

2.1 Problem 2: Logistic growth in continuous time

2.1.1 a: Finding solution of logistic equation

We have to solve the equation

$$\frac{dx}{dt} = rx(1 - \frac{x}{K}) = r(x - \frac{x^2}{K}) = \frac{rx(K-x)}{K} \tag{2.15}$$

using the separation of variables

$$\frac{K}{x(K-x)} dx = r dt \tag{2.16}$$

decomposing the right part with partial fractions

$$\frac{K}{x(K-x)} = \frac{A}{x} + \frac{B}{K-x} \tag{2.17}$$

We find A and B

$$A = \frac{K - Bx}{K - x} \tag{2.18}$$

$$B = \frac{K - A(K-x)}{x} \tag{2.19}$$

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

$$\frac{K}{x(K-x)} = \frac{1}{x} + \frac{1}{K-x} \tag{2.20}$$

Now we have to take the integral from both parts:

$$\int \frac{1}{x} + \frac{1}{K-x} dx = \int r dt \tag{2.21}$$

$$\int \frac{1}{x} + \int \frac{1}{K-x} dx = r \int dt \tag{2.22}$$

$$\ln x - \ln K - x = rt + x_0 \tag{2.23}$$

$$\ln \frac{x}{K-x} = rt + x_0 \tag{2.24}$$

$$\frac{x}{K-x} = x_0 e^{rt} \tag{2.25}$$

Finally the solution

$$x(t) = x_0 K e^{rt} \frac{1}{K + x_0(e^{rt} - 1)} \tag{2.26}$$

To prove it we explicit the computation behind our results($\lambda = r$):

$$\frac{dx}{dt} = \lambda x(1 - \frac{x}{k}) = \lambda x(\frac{k-x}{k}) \rightarrow \int \frac{k}{x(k-x)} dx = \int \lambda dt$$

$$\int [\frac{1}{x} + \frac{1}{k-x}] dx = \int \lambda dt \rightarrow \ln |x| - \ln |k-x| = \lambda t + c \rightarrow$$

$$\rightarrow \ln |k-x| - \ln |x| = -\lambda t - x \rightarrow \ln |\frac{k-x}{x}| = e^{-\lambda t - c} = \frac{k-x}{x} = \pm e^{-\lambda t} e^{-c} \text{ with } A = e^{-c}$$

$$\rightarrow \frac{k-x}{x} = Ae^{-\lambda t} = \frac{k}{x} - 1 = Ae^{-\lambda t} \rightarrow x = \frac{k}{1 + Ae^{-\lambda t}} = \frac{k}{1 + Ae^{-\lambda t}}(\frac{e^{\lambda t}}{e^{\lambda t}}) = \frac{ke^{\lambda t}}{e^{\lambda t} + A}$$

$$x_0 = x(0) = \frac{k}{1 + A} = x_0 \rightarrow k = x_0 + Ax_0 = k - x_0 = Ax_0 \rightarrow A = \frac{k - x_0}{x_0}$$

$$x = \frac{ke^{\lambda t}}{e^{\lambda t} + \frac{k - x_0}{x_0}} = \frac{kx_0 e^{\lambda t}}{e^{\lambda t} x_0 + k - x_0} = \frac{kx_0 e^{\lambda t}}{k + x_0(e^{\lambda t} - 1)}$$

2.1.2 b: Determining the stability of equilibrium points

To determine the stability of the two equilibria points found solving the equation 2.15 in 0

$$x_1^* = 0 \tag{2.27}$$
$$x_2^* = K \tag{2.28}$$

one has to derive it for x:

$$f'(x) = -r(\frac{x}{K} - 1) - \frac{rx}{K} \tag{2.29}$$

Substituting x_1^* and x_2^* into 2.29 yields:

$$f'(x_1^*) = r \text{ if } r < 0 \text{ stable equilibrium} \tag{2.30}$$
$$f'(x_2^*) = -r \text{ if } r > 0 \text{ stable equilibrium} \tag{2.31}$$

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

3.1.1 a: frequency of allele B in population

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

$$0^2 = \frac{900}{10000} = 0.09 = r^2 \rightarrow r = 0.3 \tag{3.32}$$

$$2AB = \frac{2000}{10000} = 0.20 = 2pq \rightarrow pq = 0.1 \tag{3.33}$$

$$A^2 + 2A0 = \frac{1600}{10000} = 0.16 = p^2 + 2pr \tag{3.34}$$

$$\tag{3.35}$$

$$p^2 + 2 * 0.3 * p = 0.16 \tag{3.36}$$

$$p_1 = \frac{1}{5} \tag{3.37}$$

$$p_2 = -\frac{4}{5} \text{ not acceptable} \tag{3.38}$$

$$\tag{3.39}$$

$$q = \frac{0.1}{0.2} = 0.5 \tag{3.40}$$

So, the frequency for B allele is equal to 0.5

3.1.2 b: HW Equilibrium

The equilibrium is given by $p^* = P^2 + pq + pr$

$$A^2 = \frac{1500}{10000} = 0.15 = p^2 \rightarrow p = 0.39 \tag{3.41}$$

$$p^2 + 2pr = \frac{1600}{10000} = 0.16 \rightarrow 0.15 + 2pr = 0.16 \rightarrow pr = \frac{0.01}{2} = 0.005 \tag{3.42}$$

$$2pq = \frac{2000}{10000} = 0.2 \rightarrow pq = 0.1 \tag{3.43}$$

$$\tag{3.44}$$

$$p^* = 0.15 + 0.0005 + 0.1 = 0.255$$

So, for the next round of random mating, the number of genotype $AA = (P^*)^2 * 10000 = (0.255)^2 * 10000 = 650 \neq 1500$. The first generation is not in Hardy-Weinberg equilibrium.

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:

$$\text{for } c_1 = 20 \text{ then } s_{space} = c_1^L, 20^{50} = 1.13 * 10^{65} \tag{4.45}$$

$$\text{for } c_2 = 22 \text{ then } s_{space} = c_2^L, 22^{50} = 1.32 * 10^{67} \tag{4.46}$$

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.

$$\text{for } L = 150 \text{ then } d_{space} = c^L = 4^{150} = 2.04 * 10^{90} \tag{4.47}$$

$$\text{for } L = 153 \text{ then } d_{space} = c^L = 4^{153} = 1.30 * 10^{92} \tag{4.48}$$

Comparison between the number of unique amino acid sequence and DNA unique sequences:

$$\frac{4^{150}}{20^{50}} = 1.8 * 10^{25} = (\frac{64}{20})^{50} \text{ more DNA sequences}$$

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

$$dist(A) = \frac{1}{|A|^2} \sum_{x \in A} \sum_{y \in A} d_H(x, y) \tag{5.49}$$

Hence

$$dist(F_2^n) = \frac{1}{2^{2n}} \sum_{x \in F_2^n} \sum_{y \in F_2^n} d_H(x, y) = \frac{n 2^{n-1} 2^n}{2^{2n}} = \frac{n}{2} \tag{5.50}$$

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 $q = 4$

$$Ed_H(X, Y) = \frac{n(q-1)}{q} = \frac{L(4-1)}{4} = \frac{3}{4} L \tag{5.51}$$

So, for binary sequences we can compute the average distance with equation 5.51 as well.

5.1.2 b: Number of sequences at a certain distance

With given sequences of length L for Hamming distance K and alphabet A:

$$c(L, K) = (A - 1)^K \frac{L!}{(L - K)!K!}$$

For a binary sequence ($A = 2$) of length L at distance $K = 1$:

$$c(L, K) = (2 - 1)^K \frac{L!}{(L - K)!K!} \text{ for } K \leq L$$

$$c(L, K) = (2 - 1)^1 \frac{L!}{(L - 1)!1!} = \frac{L!}{(L - 1)!} = L$$

For DNA sequences ($A = 4$) of length L at distance $K = 1$:

$$c(L, K) = (4 - 1)^K \frac{L!}{(L - K)!K!}$$

$$c(L, K) = 3^K \frac{L!}{(L - K)!K!} \text{ for } K \leq L$$

$$c(L, K) = 3^1 \frac{L!}{(L - 1)!1!} = 3L$$

5.1.3 c: Replication of sequences

Considering H_{ij} as d and q_{ij} as q :

$$q_{ij} = p^d (1 - p)^{L-d}$$

$$\ln q = \ln(p^d (1 - p)^{L-d}) = \ln p^d + \ln(1 - p)^{L-d}$$
$$d \ln p + (l - d) \ln(1 - p) = d \ln p - l \ln(1 - p) - d \ln(1 - p)$$
$$d(\ln p - \ln(1 - p)) = \ln q - l \ln(1 - p)$$
$$d = \frac{\ln q - l \ln(1 - p)}{\ln p - \ln(1 - p)}$$

6.1 Problem 6: Quasispecies

6.1.1 a: Find mutation-selection matrix w

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

If Q is error free then $Q = I$ but here Q is not only error free but also the genotypes are replicated with probability q. So,

$$Q = (q_{ij}) = (f_j q_{ji}) = \begin{bmatrix} q & 1 - q \\ 1 - q & q \end{bmatrix}$$

with $f_1 = 1$ and $f_0 > 1 \rightarrow W = (w_{ij}) = (f_j q_{ji}) = \begin{bmatrix} w_{00} & w_{01} \\ w_{10} & w_{11} \end{bmatrix} = \begin{bmatrix} f_0 q_{00} & f_1 q_{01} \\ f_0 q_{10} & f_1 q_{11} \end{bmatrix} = \begin{bmatrix} f_0 q & 1-q \\ f_0(1-q) & q \end{bmatrix}$

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

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

$$\begin{bmatrix} f_0 q - \lambda & 1-q \\ f_0 - f_0 q & q - \lambda \end{bmatrix} = 0 \rightarrow (f_0 q - \lambda)(q - \lambda) - (f_0 - f_0 q)(1-q) \rightarrow$$

$$f_0 q^2 - f_0 q \lambda - \lambda q + \lambda^2 - f_0 + f_0 q + f_0 q - f_0 q^2 = 0$$
$$\lambda^2 - (f_0 q + q)\lambda + (2 f_0 q - f_0) = 0$$
$$\begin{cases} \lambda_1 = \frac{(f_0 q + q) + \sqrt{(f_0 q + q)^2 - 4(2 f_0 q - f_0)}}{2} \\ \lambda_2 = \frac{(f_0 q + q) - \sqrt{(f_0 q + q)^2 - 4(2 f_0 q - f_0)}}{2} \end{cases}$$

6.1.2   b: Finding nontrivial solutions

Φ is the largest eigen value of, so $\Phi = \lambda_1$ **nontrivial solution**.

6.1.3   c: Equilibrium point for $f_0 = f_1 = 1$

if $f_0 = f_1 = 1$ then

$$\Phi = \frac{2q + \sqrt{(2q)^2 - q(2q - 1)}}{2} = \frac{2q + \sqrt{4q^2 - 4(2q - 1)}}{2} = \frac{2q + 2\sqrt{q^2 - 2q + 1}}{2} = q + \sqrt{(q-1)^2} = q + (q - 1)$$

$$\begin{cases} \Phi_1 = 2q - 1 \\ \Phi_2 = 1 \end{cases}$$

$$\begin{pmatrix} x_0' \\ x_1' \end{pmatrix} = \begin{pmatrix} q & 1-q \\ 1-q & q \end{pmatrix} \begin{pmatrix} x_0 \\ x_1 \end{pmatrix} - \Phi \begin{pmatrix} x_0 \\ x_1 \end{pmatrix}$$

if $\Phi = 1$, then:

$$\begin{pmatrix} x_0' \\ x_1' \end{pmatrix} = \begin{pmatrix} qx_0 + (1-q)x_1 \\ (1-q)x_0 + qx_1 \end{pmatrix} - \begin{pmatrix} x_0 \\ x_1 \end{pmatrix}$$

$$x_0' = qx_0 + (1-q)x_1 - x_0 = f(x_0, x_1)$$
$$x_1' = (1-q)x_0 + qx_1 - x_1 = g(x_0, x_1)$$

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

$$qx_0^* + x_1^* - qx_1^* - x_0^* = 0 \rightarrow (1-q)x_0^* = (1-q)x_1^*$$
$$1 - q)x_0^* + qx_1^* - x_1^* = 0 \rightarrow (1-q)x_0^* = (1-q)x_1^*$$
$$\rightarrow x_0^* = \frac{1-q}{1-q}x_1^* = x_1^* \rightarrow x_0^* = x_1^*$$

Population type 0 is equal to population type $1 = \frac{1}{2}$

6.1.4 d

$f_0 \geq 0$ and $q = 1$

$$q = 1 \rightarrow \Phi = \frac{(f_0 + 1) + \sqrt{(f_9 + 1)^2 - 4(2f_0 - f_0)}}{2} = \frac{(f_0 + 1) + \sqrt{(f_0 + 1)^2 - 4f_0}}{2} = \frac{(f_0 + 1) + (f_0 - 1)}{2}$$

$$\begin{cases} \Phi_1 = \frac{f_0 + 1 + f_0 - 1}{2} = f_0 \\ \Phi_2 = \frac{f_0 + 1 - f_0 + 1}{2} = 1 \end{cases}$$

Since Φ should be the largest and $f_0 \gg 1$ we choose Φ_1.

$$\begin{pmatrix} x_0' \\ x_1' \end{pmatrix} = \begin{pmatrix} q & 1 - q \\ 1 - q & q \end{pmatrix} \begin{pmatrix} x_0 \\ x_1 \end{pmatrix} - f_0 \begin{pmatrix} x_0 \\ x_1 \end{pmatrix} = \begin{pmatrix} q & 1 - q \\ 1 - q & q \end{pmatrix} \begin{pmatrix} x_0 \\ x_1 \end{pmatrix} - \begin{pmatrix} f_0 x_0 \\ f_0 x_1 \end{pmatrix}$$

$$x_0' = qx_0 + (1 - q)x_1 - f_0 x_0 = f(x_0, x_1)$$

$$x_1' = (1 - q)x_0 + qx_1 - f_0 x_1 = g(x_0, x_1)$$

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

$$qx_0^* + x_1^* - qx_1^* - f_0 x_0^* = 0 \rightarrow (f_0 - q)x_0^* = (1 - q)x_1^*$$

$$1 - q)x_0^* + qx_1^* - f_0 x_1^* = 0 \rightarrow (f_0 - q)x_0^* = (f_0 - q)x_1^*$$

$$\rightarrow x_0^* = \frac{f_0 - q}{1 - q} = \frac{f_0 - 1}{0} \rightarrow x_0^* = \infty \text{ whole the population is type 1}$$

$$\rightarrow x_1^* = \frac{1 - q}{f_0 - q} = \frac{1 - 1}{f_0 - 1} \rightarrow x_1^* = 0 \text{ no type 0 in the population}$$