

Evolutionary Dynamics

# Assignment #01

Aliaksandr Yudzin, Lorenzo Gatti

Wednesday 3<sup>rd</sup> October, 2012

## 1.1 Problem 1: Logistic difference equation

### 1.1.1 a: Finding the equilibrium points

We can find equilibrium points from

$$f(x, a) = ax(1 - x) = x \quad (1.1)$$

as the roots of the quadratic equation:

$$ax^2 - (a - 1)x = x(ax - (a - 1)) = 0 \quad (1.2)$$

which are:

$$x_1^* = 0 \quad (1.3)$$

$$x_2^* = \frac{(a - 1)}{a} \quad (1.4)$$

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  $x^*$ :

$$f' = -ax - a(x - 1) = a - 2ax \quad (1.5)$$

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  $a = 0.9$

$$x_1^* = 0 \rightarrow |a| = |0.9| < 1 \text{ attractive} \quad (1.6)$$

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

$$(1.8)$$

When  $a = 2.1$

$$x_1^* = 0 \rightarrow |a| = |2.1| > 1 \text{ not attractive} \quad (1.9)$$

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

$$(1.11)$$

When  $a = 3.58$

$$x_1^* = 0 \rightarrow |a| = |3.58| > 1 \text{ not attractive} \quad (1.12)$$

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

$$(1.14)$$

In a conceptual view we can see for  $a$  values in excess of 3.57, the orbits  $x(t, x_0) = x_0, x_1, x_2, \dots$  depend crucially on the initial condition  $x_0$ . Slight variations in  $x_0$  result in dramatically different orbits, an important characteristic of chaos.

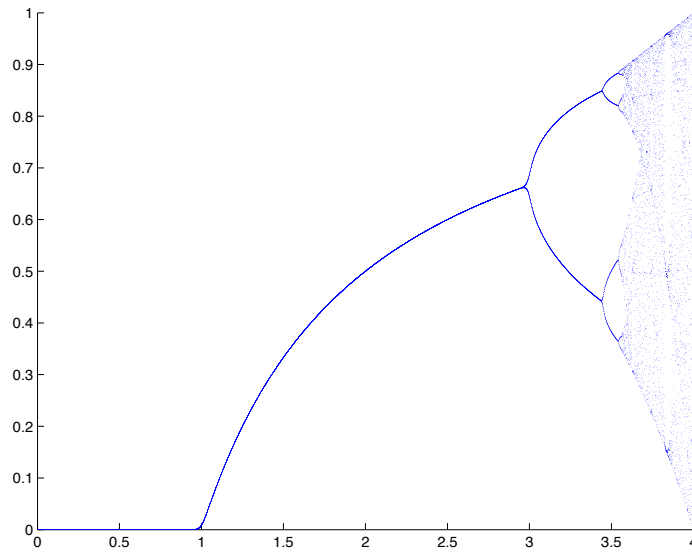


Figure 1.1: Logistic Map Bifurcation Diagram

## 2.1 Problem 2: Logistic growth in continuous time

### 2.1.1 a: Finding solution of logistic equation

We have to solve the equation

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

using the separation of variables

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

decomposing the right part with partial fractions

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

We find A and B

$$A = \frac{K - Bx}{K - x} \quad (2.18)$$

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

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

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

Now we have to take the integral from both parts:

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

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

$$\ln x - \ln K - x = rt + x_0 \quad (2.23)$$

$$\ln \frac{x}{K-x} = rt + x_0 \quad (2.24)$$

$$\frac{x}{K-x} = x_0 e^{rt} \quad (2.25)$$

Finally the solution

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

### 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

$$x_1^* = 0 \quad (2.27)$$

$$x_2^* = K \quad (2.28)$$

one has to derive it for  $x$ :

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

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

$$f'(x_1^*) = r \text{ if } r < 0 \text{ stable equilibrium} \quad (2.30)$$

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

One finds that if  $f'(x_1)$  with  $x_1 = 0$ , then  $x^*$  is always stable, while if  $f'(x_2) > 0$  with  $x_2 = K$  is repelling.

### 3.1 Problem 3: Hardy-Weinberg equilibrium

#### 3.1.1 a: frequency of allele B in population

Assuming HWE for three alleles (allele and genotype frequencies):

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

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 frequencies

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

$$2AB = \frac{2000}{10000} = 0.20 = 2pq \rightarrow pq = 0.1 \quad (3.33)$$

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

$$(3.35)$$

$$p^2 + 2 * 0.3 * p = 0.16 \quad (3.36)$$

$$p_1 = \frac{1}{5} \quad (3.37)$$

$$p_2 = -\frac{4}{5} \text{ not acceptable} \quad (3.38)$$

$$(3.39)$$

$$q = \frac{0.1}{0.2} = 0.5 \quad (3.40)$$

So, the frequency for B allele is equal to 0.5

#### 3.1.2 b: HW Equilibrium

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

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:

$$\text{for } c_1 = 20 \text{ then } s_{space} = c_1^L, 20^{50} = 1.13 * 10^{65} \quad (4.42)$$

$$\text{for } c_2 = 22 \text{ then } s_{space} = c_2^L, 22^{50} = 1.32 * 10^{67} \quad (4.43)$$

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.

$$\text{for } L = 150 \text{ then } d_{space} = c^L = 4^{150} = 2.04 * 10^{90} \quad (4.44)$$

$$\text{for } L = 153 \text{ then } d_{space} = c^L = 4^{153} = 1.30 * 10^{92} \quad (4.45)$$

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

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

### 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  $x = (x_1, \dots, x_n)$  and  $y = (y_1, \dots, y_n)$  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

$$dist(A) = \frac{1}{|A|^2} \sum_{x \in A} \sum_{y \in A} d_H(x, y) \quad (5.46)$$

Hence

$$dist(F_2^n) = \frac{1}{2^{2n}} \sum_{x \in F_2^n} \sum_{y \in F_2^n} d_H(x, y) = \frac{n2^{n-1}2^n}{2^{2n}} = \frac{n}{2} \quad (5.47)$$

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  $2 \leq q \leq 4$ , (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  $q = 4$

$$Ed_H(X, Y) = \frac{n(q-1)}{q} = \frac{L(4-1)}{4} = \frac{3}{4}L \quad (5.48)$$

So, for binary sequences we can compute the average distance with equation 5.48 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  $q$  :

$$\begin{aligned}
 q_{ij} &= p^d (1-p)^{L-d} \\
 \ln q &= \ln(p^d (1-p)^{L-d}) = \ln p^d + \ln(1-p)^{L-d} \\
 d \ln p + (l-d) \ln(1-p) &= d \ln p - l \ln(1-p) - d \ln(1-p) \\
 d(\ln p - \ln(1-p)) &= \ln q - l \ln(1-p) \\
 d &= \frac{\ln q - l \ln(1-p)}{\ln p - \ln(1-p)}
 \end{aligned}$$

## 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  $q$ .

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

$\Phi$  is the largest eigen value of  $w$ , while  $x$  is the eigen vector of  $w$ .

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

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

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

$$\begin{pmatrix} x'_0 \\ x'_1 \end{pmatrix} = \begin{pmatrix} 0 & q \\ f q & 0 \end{pmatrix} \begin{pmatrix} x_0 \\ x_1 \end{pmatrix} - q\sqrt{f} \begin{pmatrix} x_0 \\ x_1 \end{pmatrix} \rightarrow \begin{pmatrix} x'_0 \\ x'_1 \end{pmatrix} = \begin{pmatrix} q x_1 \\ f q x_0 \end{pmatrix} + \begin{pmatrix} -q\sqrt{f} x_0 \\ -q\sqrt{f} x_1 \end{pmatrix} = \begin{pmatrix} q x_1 - q\sqrt{f} x_0 \\ f q x_0 - q\sqrt{f} x_1 \end{pmatrix}$$

$$x'_0 = q x_1 - q\sqrt{f} x_0 = f(x_0, x_1) \quad (6.49)$$

$$x'_1 = f q x_0 - q\sqrt{f} x_1 = g(x_0, x_1) \quad (6.50)$$

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

$$\begin{cases} x_0^* = \frac{1}{\sqrt{f}} x_1^* \\ x_0^* = \frac{\sqrt{f}}{f} x_1^* = \frac{1}{\sqrt{f}} x_1^* \end{cases} \rightarrow x_0^* = \frac{1}{\sqrt{f}} x_1^*$$

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

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

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 genotype 0 and 1 is the same in whole population at steady state.

### 6.1.2 d

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

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