

Evolutionary Dynamics

Assignment #01

Aliaksandr Yudzin, Lorenzo Gatti

Tuesday 2nd October, 2012

1.1 Problem 1: Logistic difference equation

We can find stable 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^* :

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

Substituting x_1 and x_2 into 1.5 yields:

$$f'(x_1) = a \quad (1.6)$$

$$f'(x_2) = 2 - a \quad (1.7)$$

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 $x(t, x_0) = x_0, x_1, x_2, \dots$ depend crucially on the initial condition x_0 . Slight variations in x_0 result in dramatically different orbits, an important characteristic of chaos.

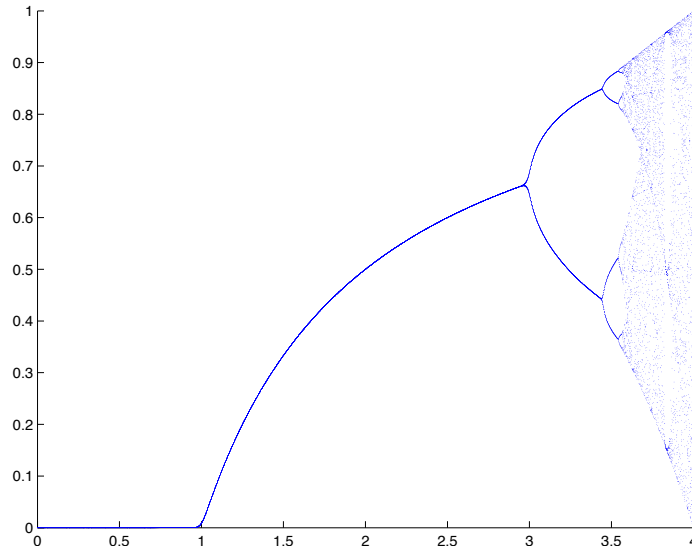


Figure 1.1: Logistic Map Bifurcation Diagram

2.1 Problem 2: Logistic growth in continuous time

We have to solve the equation

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

using the separation of variables

$$\frac{K}{x(K-x)} dx = r dt \quad (2.9)$$

decomposing the right part with partial fractions

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

We find A and B

$$A = \frac{K - Bx}{K - x} \quad (2.11)$$

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

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

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

Now we have to take the integral from both parts:

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

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

$$\ln x - \ln K - x = rt + x_0 \quad (2.16)$$

$$\ln \frac{x}{K-x} = rt + x_0 \quad (2.17)$$

$$\frac{x}{K-x} = x_0 e^{rt} \quad (2.18)$$

Finally the solution

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

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

$$x_1 = 0 \quad (2.20)$$

$$x_2 = K \quad (2.21)$$

one has to derive it for x :

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

Substituting x_1 and x_2 into 2.22 yields:

$$f'(x_1) = -1 \quad (2.23)$$

$$f'(x_2) = -r \quad (2.24)$$

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

3.1 Problem 3: Hardy-Weinberg equilibrium

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

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

$$2AB = \frac{2000}{10000} = 0.20 = 2pq \rightarrow pq = 0.1 \quad (3.26)$$

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

$$(3.28)$$

$$p^2 + 2 * 0.3 * p = 0.16 \quad (3.29)$$

$$p_1 = \frac{1}{5} \quad (3.30)$$

$$p_2 = -\frac{4}{5} \text{ not acceptable} \quad (3.31)$$

$$(3.32)$$

$$q = \frac{0.1}{0.2} = 0.5 \quad (3.33)$$

So, the frequency for B allele is equal to 0.5

3.1.1 HW Equilibrium

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

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 n we consider there are:

$$\text{for } c = 20 \text{ then } s_{space} = c^L, 20^{50} = 1.13 * 10^{65} \quad (4.35)$$

$$\text{for } c = 22 \text{ then } s_{space} = c^L, 22^{50} = 1.32 * 10^{67} \quad (4.36)$$

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.

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

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

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 $x = (x_1, \dots, x_n)$ 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

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

Hence

$$\text{dist}(F_2^n) = \frac{1}{2^{2n}} \sum_{x \in F_2^n} \sum_{y \in F_2^n} d_H(x, y) = \frac{n2^{n-1}2^n}{2^{2n}} = \frac{n}{2} \quad (5.40)$$

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

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

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

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

6.1 Problem 6: Quasispecies