ETH zürich D BSSE Computational Biology Group

Evolutionary Dynamics

Exercises 2

Prof. Dr. Niko Beerenwinkel
Dr. Katharina Jahn
Dr. Rob Noble
3rd October 2019

Problem 1: Sequence space

Consider the sequence space of amino acids of length $L = 20$ and that of DNA sequences encoding for them,

(a) How many unique amino acid sequences are there? **(0.5 point)**

(b) What is the length of a DNA sequence coding for $L = 20$ amino acids? How many unique DNA sequences exists of that length? **(0.5 point)**

Problem 2: Hamming distance

Consider an alphabet $\mathcal{A}$ of size $|\mathcal{A}| = A$. For a binary alphabet, one has $\mathcal{A} = \{0,1\}$ and $A = 2$, and for DNA, one has $\mathcal{A} = \{A,T,C,G\}$ and $A = 4$. We are studying sequences $S \in \mathcal{A}^L$ of length L. Assume sequences are random with a uniform distribution,

(a) What is the average Hamming distance between two random binary sequences? What is the expected Hamming distance for two random DNA sequences? **(1 point)**

(b) Given a binary sequence of length L, how many sequences exist at a Hamming distance two from it? How many at distance K with $K \leq L$? Repeat the calculation for DNA sequences. **(2 points)**

Problem 3: Quasispecies

Consider the quasispecies equation with two genotypes $0, 1$ (i.e., binary sequences of length 1). Let the fitness of genotype 0 be $f_0 > 1$, and the fitness of genotype 1 be $f_1 = 1$. Moreover, genotypes are replicated error-free with probability q,

(a) Write down the mutation-selection matrix W and find its two eigenvalues. **(2 points)**

(b) To which eigenvalue corresponds the non-trivial equilibrium point? **(1 point)**
Hint: Revise Perron-Frobenius theorem.

(c) Examine the dynamics of the quasispecies equations and confirm results obtained in (b). *Hint*: This is a programming exercise. **(1 point)**

(d) What is the equilibrium point for $f_0 = f_1 = 1$? **(1 point)**

(e) Calculate the equilibrium point in the limit of low mutation rate ($q \approx 1$). **(1 point)**