

Home assignment #2 – computational models in genetics and living systems

Fitness landscapes

1. (45 points) **NK model:** Simulate the NK model. Take $N=14$ and use 3 different values of K ($K=0, 4, 10$). Use a binary alphabet. Every binary vector of length N (g_1, g_2, \dots, g_N) then represents a genotype. Define the neighborhood between genotypes such that genotypes that differ in exactly one position are nearest neighbors. Define the fitness value of each genotypes as follows:

The fitness of the genotype is the sum of local fitness values f_i , where f_i depends on the identity of the $(K+1)$ -tuple (g_i, \dots, g_{i+K}) – use cyclic indices if $i+K > N$. Draw the values of f_i from a uniform distribution.

For example, if $K=1, N=5$ and the vector is 00101, its fitness is $f(00)+f(01)+f(10)+f(01)+f(10)$. Draw the fitness values for all binary pairs from a uniform distribution.

- i) (20 points) Calculate the spatial correlation for each of the fitness landscapes, as follows: choose an arbitrary starting point and then create a random walk on the landscape by randomly choosing one of the nearest neighbors of the current genotype. Create trajectories of length 2^N (does the trajectory length matter?). Then calculate the autocorrelation of the vector of fitness values along the trajectory. Does the correlation vary with K ? How? Can you explain this trend?
For definition and calculation of autocorrelation – check the additional help page.
- ii) (10) How many local maxima does each fitness landscape contain? A local maximum is a genotype for which all the nearest neighbors have lower fitness.
- iii) (15) Construct trajectories of non-decreasing fitness: again start from an arbitrary genotype. Now choose each time its nearest neighbor that has the highest fitness. If multiple neighbors have equal fitness choose one of them randomly. If the highest fitness among neighbors is equal to the current genotype's fitness choose one of these equal fitness neighbors. Stop when no such neighbor exists anymore. Document the length of the trajectory. Repeat this for all genotypes in the landscapes as starting points. Plot the distribution of trajectory lengths. Again, does it depend on K ? why?

2. Uncorrelated fitness landscape:

- i) (35) Create a fitness landscape of 2^N ($N=10$) genotypes and randomly choose the fitness rank for all the genotypes. The actual fitness values are not necessary here. Define the neighborhood between genotypes such that genotypes that differ in only one position are nearest neighbors.
Calculate the correlation (10), number of local maxima (10) and the distribution of trajectory lengths (10) for this landscape. Compare them to the NK landscape (5).

- ii) (10) Bonus: estimate the number of local maxima in an uncorrelated landscape from simple considerations as a function of N . **Hint:** if you choose an arbitrary genotype in an uncorrelated landscape, what is the probability that it is a local maximum?
You can simulate uncorrelated landscapes with different values of N to check your answer.

3. Genetic load

(20 points) Calculate the genetic load of a population under single-locus selection assuming over-dominance (namely, the heterozygous genotype has the highest fitness). Assume equilibrium and Use the expressions for genotype frequencies at equilibrium.