Environmental Barriers and Population Fitness

Jason Rohrer

Created: March 19, 2002; Last modified: March 19, 2002

1 Introduction

The goal of this work was to determine how environmental barriers to free crossbreeding affect overall population fitness. Our hypothesis was that populations that can freely interbreed will achieve less fitness than populations that are segmented by the environment into regions of free breeding with only rare breeding between regions.

To test this hypothesis, we set up a computer simulation that featured a population of individuals in an environment that discouraged free movement and interbreeding. The majority of this paper describes various features of our simulation.

The final section presents the results we obtained from our simulation. From these experiments, it seems that environmental barriers do improve overall population fitness.

2 Environment

Individuals exist on a fractal ($1/f$) height map where the cost of traveling between region increases as the height disparity between the regions increases.

Assume that we are given a two-dimensional white noise field function that has the form

$$w : (x, y) \to [-1, 1].$$

Such a function maps each point on the real plane to a noise value. Assume that this noise field is *not* scale-invariant, in other words

$$\lim_{h \to 0} w(x + h, y + h) = w(x, y).$$

Thus, the frequency information in our noise field decreases as the scale decreases. We can use such a field to produce a fractal noise function as follows:

$$g(x, y) = z \sum_{f=1}^{c} \frac{w(x \cdot f, y \cdot f)}{f},$$

where c is the maximum frequency to include in the sum, and z is the following normalization factor:

$$z = \frac{1}{\sum_{f=1}^{c} \frac{1}{f}}.$$

Intuitively, we are extracting information from w at various scales and summing it together, where each scale's contribution is weighted by the inverse of its frequency. We give low frequency information the most weight and high frequency information the least weight. We can think of the low frequency information as our "base" terrain and the higher frequency information as perturbing that base.

3 Movement distribution

Individuals move randomly throughout the environment. When on a surface with no gradient (a level surface), an individual will move in all directions with equal probability. On a surface with a non-zero gradient, the likelihood of a move is smaller in the directions that are closer to parallel with the gradient. Moves closer to perpendicular to the gradient are made with higher probability. More formally, an individual's move in a given timestep is chosen from a distribution over all length-one direction vectors $(\Delta x, \Delta y)$ with density proportional to the inverse exponential of the squared surface rise in that direction. With a height function $h : (x, y) \to \Re$ and at a given point $P_0 = (x_0, y_0)$, we get the density function:

$$p(\Delta x, \Delta y|\alpha, P_0) = c \cdot \exp\left(-\alpha\left[\Delta x\left(\frac{\delta h}{\delta x}\right)_{P_0} + \Delta y\left(\frac{\delta h}{\delta y}\right)_{P_0}\right]^2\right), \text{ if } (\Delta x)^2 + (\Delta y)^2 = 1,$$

and 0 otherwise, where α is a parameter that controls how much the slope affects movement. If we let β represent our partial height derivative with respect to x and λ represent our partial with respect to y, we get:

$$p(\Delta x, \Delta y|\alpha, \beta, \lambda) = c \cdot \exp\left(-\alpha(\beta\Delta x + \lambda\Delta y)^2\right), \text{ if } (\Delta x)^2 + (\Delta y)^2 = 1.$$

There is no closed-form solution for the cumulative distribution function (CDF) of this distribution, even if we look at the marginals in isolation, so we must employ a numerical technique to sample from it. We use the numerical technique of rejection sampling.[1]

To derive a valid approximation function, we need to know the maximum of $p(\Delta x, \Delta y|\alpha, \beta, \lambda)$ for Δx and Δy on the unit circle, our interval of interest. Using Maple's `maximize` function, we can determine that for any values of β and λ, for any $\alpha \geq 0$, and for *any* values of Δx and Δy (even off of the unit circle), the density never rises above 1.

We choose the following rather naive approximation function for rejection sampling

$$g(\Delta x, \Delta y) = \begin{cases} 1, & \text{if } (\Delta x)^2 + (\Delta y)^2 = 1 \\ 0, & \text{otherwise.} \end{cases}$$

Perhaps a more accurate approximation function would be possible, but we do not derive one here, since our function is complex enough to make such derivation difficult. See Figure 1 on page 3 for a sample plot of this distribution. Note that with rejection sampling, the accuracy of the approximation function only effects the efficiency, not the accuracy, of the sampling procedure.

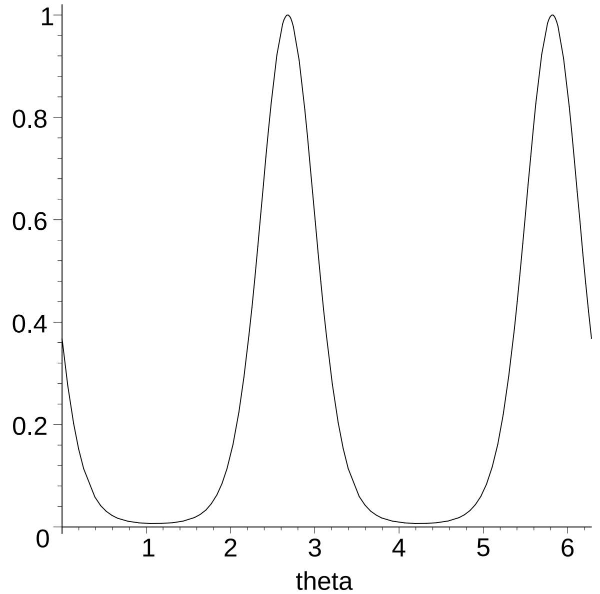

Figure 1: The movement distribution for angle θ around the unit circle with $\alpha = 1$, $\beta = 1$, and $\lambda = 2$.

4 Mating distributions

The fitness of an individual can be determined according to any measure. An interface is provided that allows for arbitrary measures to be plugged in.

At the end of each epoch of timesteps, an individual may mate with other individuals in the environment and produce offspring. The chance of an individual i mating is proportional to its fitness, f_i, according to the following distribution:

$$p_{mate}(i) = \frac{f_i}{\sum_{j=1}^{n} f_j}.$$

The number of matings per epoch is controlled by the parameter γ. During an epoch, γn individuals are drawn (with repeats allowed) according to p_{mate}. Thus, an individual may mate more than once in an epoch.

After an individual i is chosen to mate, a mate is chosen for that individual from a distribution with density increasing as the inverse of the distance between the individuals. For individual j separated from the chosen individual by distance D_{ij}, we have

$$p_i(j) = \eta \cdot \exp\left(-\eta D_{ij}\right).$$

Note that the above distribution formula is normalized as if it were a continuous distribution, while our actual random variable J is discrete. By applying rejection sampling techniques to this distribution (as opposed to the standard inverse CDF techniques for an exponential distribution), we effectively renormalize the distribution so that it sums to one over our discrete variable space. Also note that this distribution is over the *entire* population (not just those selected for mating). Thus, an unfit individual might still participate in a mating if it is close to a fit individual.

The produced offspring is placed into the environment halfway between the two parent individuals. The population is capped by a threshold. When the threshold is crossed, individuals are selected uniformly at random from the population and removed. Thus, an individual's survival is unaffected by its fitness: fitness is rewarded only with the production of offspring.

5 Experiments

For these experiments, we constructed a fitness measure based on a fractal fitness surface. This surface was similar to, but separate from, the fractal surface that comprised the environment Each individual was represented by a point on this surface, and the fitness of an individual was determined by the height of the individual's point on the fitness surface. Individuals could traverse the environment surface, as described throughout this paper, but they could not move about on the fitness surface. When two individuals mated, their offspring was placed directly between them on the fitness surface. A mutation took the form of adding a random value to one of the coordinates of a fitness surface point. Thus, our fitness function was replete with local maxima. We hypothesized that environmental barriers would help the population as a whole avoid these local maxima and discover the global maximum.

We performed 80 separate runs of 100 generation each, computing averaged statistics (maximum fitness, mean fitness, and minimum fitness) for each run. We used a different η value for each run, testing 80 linearly spaced η settings in the range $[0.001, 20]$. Larger η values correspond to larger environmental barriers to free mating. The smallest η value, 0.001, can be thought of as our control case: environmental barriers were negligible, and the population crossbred freely across the entire environment.

6 Results

For all three averaged statistics, we see a positive correlation between increasing η (increasing the effects of environmental barriers) and an increase in population fitness. Thus, the results of these experiments support our hypothesis. Scatter plots for our results, with best-fit lines, can be seen in Figures 2, 3, and 4.

References

[1] A. Gelman, J. Carlin, H. Stern, and D. Rubin, *Bayesian Data Analysis.* Chapman and Hall, 2000.

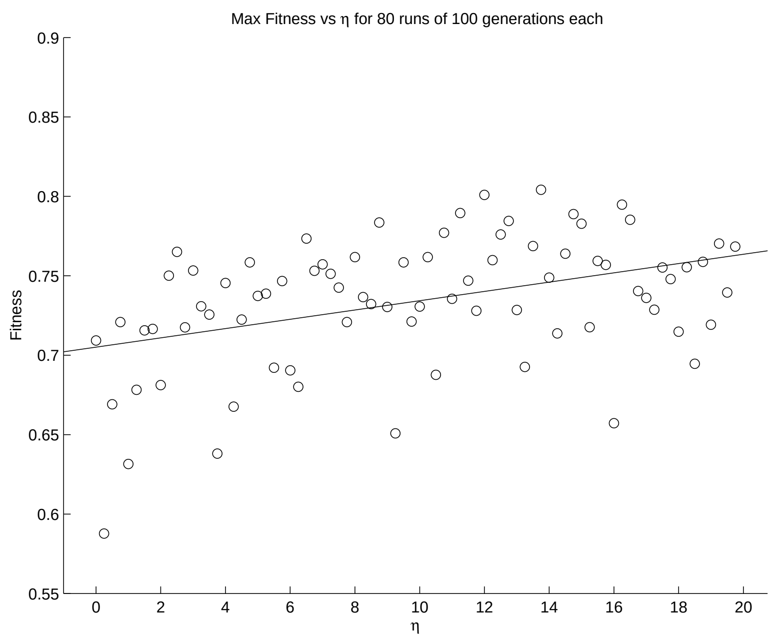

Figure 2: A scatter plot of the population maximum fitness averaged over 100 generations using each of 80 different η settings in the range $[0.001, 20]$, with $\alpha = 1.0$, $\gamma = 0.125$, and a mutation rate of 0.01.

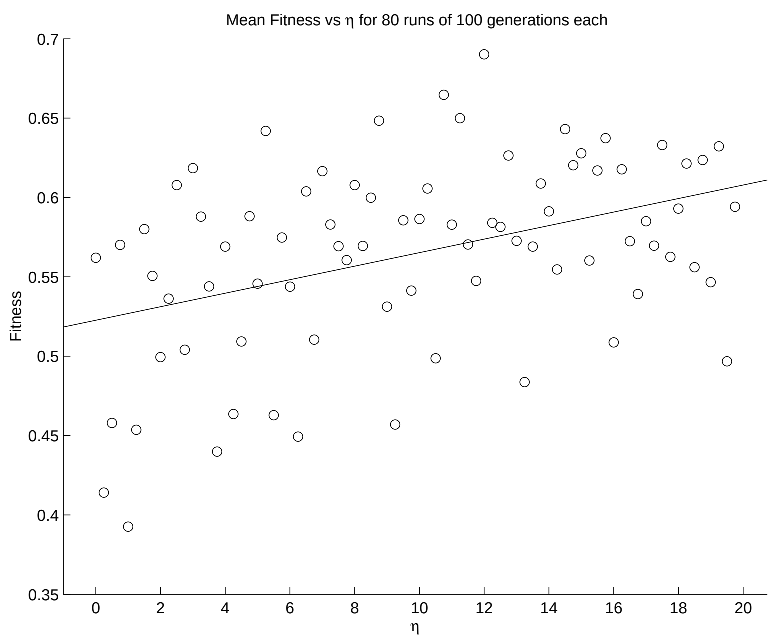

Figure 3: A scatter plot of the population mean fitness averaged over 100 generations using each of 80 different η settings in the range $[0.001, 20]$, with $\alpha = 1.0$, $\gamma = 0.125$, and a mutation rate of 0.01.

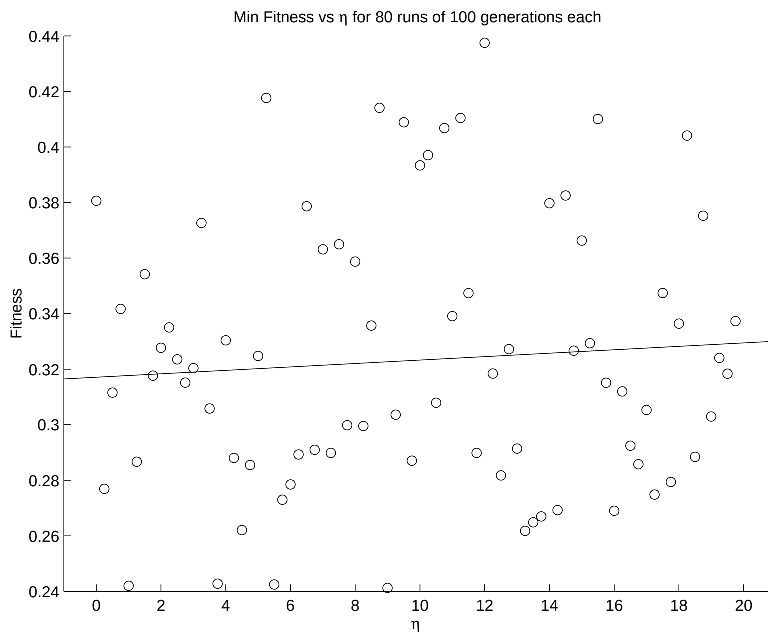

Figure 4: A scatter plot of the population minimum fitness averaged over 100 generations using each of 80 different η settings in the range $[0.001, 20]$, with $\alpha = 1.0$, $\gamma = 0.125$, and a mutation rate of 0.01.